数学世界的探奇之旅

The Uncanny Relationship of Mathematics and the Physical World

[英] 布赖恩·克莱格（Brian Clegg） 著
胡小锐 译

品读数学与人类的浪漫故事

在丰富多彩的现实世界中，探寻数学的秘密和魔力，

ARE NUMBERS REAL?

中信出版集团

版权信息

书名:数学世界的探奇之旅

作者:[英]布赖恩·克莱格

译者:胡小锐

ISBN:9787508680330

中信出版集团制作发行

献给

吉莲、丽贝卡和切尔西

当把数学原理应用于现实时，它们是不确定的；当它们确定时，又不适用于现实。

——阿尔伯特·爱因斯坦，《相对论杂谈》（1922）

第1章
虚拟的“居民”?

Are Numbers Real?

本书将探讨一个非常重要的问题：数字，或者说广义的数学，是一种真实的存在吗？这个问题对于科学家乃至我们所有人而言都具有非常重要的意义。然而，包括科学家在内的大多数人几乎从来不会考虑这个问题。

乍一看，这似乎是一个非常荒唐的问题，不要说写成一本书，就是考虑30秒钟的时间都是一种浪费。数字当然是一种真实的存在。看一眼银行对账单，就能找到答案，因为那上面有大量的数字，其中大多数都是负数，表示从账户流出的现金。数学专业人士也知道这个问题的答案，因为我们上学时要完成大量作业，接触的都是真实的数字。但是，我在这里所说的“真实”具有不同的含义。加深对科学的理解，确定数字和数学是否真实地存在于宇宙之中，具有非常重要的意义。在人们没有想到数字前，它们也是存在的吗？还是说数字是人类的重要发明，是一个重要的虚幻世界中的虚拟“居民”？

我们知道，数学可以完全脱离现实，与现实世界不发生任何本质上的联系。事实上，数学家一直是这样做的。数学（mathematics或maths）的最终目的就是制定一系列规则，帮助我们完成整个过程并最

终得出某个结果。在定义这些规则时，我们可以使其与我们对现实世界的观察结果保持一致，也可以根据我们的意愿，让它与现实世界截然不同，以产生怪诞而美妙的效果。有的数学家就喜欢这种遨游另类世界的梦幻之旅。

举一个简单的例子。现实世界有三个空间维度（然而，物理学的弦理论试图结合考虑重力等自然力，要求空间具有9个或10个维度），但是空间有1、2、4、79个还是5 000个维度，对于数学而言都不是问题。“大魔群”这个数学概念深受数学界的欢迎，该群的元素是在196 883维空间中旋转。在研究“大魔群”时，我们难免会想起《绿野仙踪》中那句非常经典的话：“托托，我觉得我们现在已经不在堪萨斯了。”

尽管数学家在研究这些群时，考虑的都是纽结这种非常普通的内容，但是他们对纽结的定义与我们系鞋带时打的绳结没有任何相似之处。出于便利实用的考虑，这些数学家规定他们用来打结的“绳索”必须头尾相接，形成一个不间断的环。我们知道现实世界的绳结是不一样的，虽然数学家（他们不是非常关注世俗生活）也知道两者之间的不同，但是他们并不在乎，因为他们有权制定这样的规则。

同样的道理，我们也可以设计一个数学系统，并规定2 + 2 = 5。这个等式在现实世界行不通，但是在设计一个数字系统时，完全可以给出这样的定义。数学上有一个常用的系统，叫作时钟算术，可以将2 + 2的值定义为0或者1。在这个系统中，数字并不是一直累加，而是像钟表上的数字一样，达到某个特定值就会重新变为0。的确，这些数字与现实世界存在某种相似性。从名称可以看出，时钟算术与时钟有相似之处。例如，在12小时制的钟面上，9 + 6 = 3。与传统的计数方式相比，这种运算可以更好地表示循环变化。这个例子说明两个问题：第一，数学具有任意性；第二，下定义时必须谨慎。尽管钟表上的数字9与数山羊时的数字9有某些相同之处，但两者是不一样的。

反过来，从现实世界这个角度来看，即使没有丰富的数学知识，我们也能走好自己的人生旅程。长久以来，绝大多数人都是这样生活的。一些非常简单的算术就像是预先编写的程序一样。比如，在碗中放入一个东西，然后再放一个，又悄悄地把第二个藏回手心。这时候，无论是狗还是小孩，在看到碗中只有一个东西时，都会感到惊讶。对于哺乳动物而言，“1 + 1 = 2”似乎是一个初级的程序。毫无疑问，这也是一个非常重要的程序，因为在面对超过一个敌人时，它可以帮助你计算取胜的可能性。其余的数学知识，大多是后来我们不断获取的。然而，数学的重要意义已经得到了证明。

如果没有数学，在现代文明中占有重要地位的科学与技术几乎都会化为泡影。数学贯穿于我们的生活当中，我们不仅在日常活动（如商店中发生的交易）中要用到数学，在了解疾病分布或者选举结果的影响力时也离不开数学。数学是一门非常重要的学科，它可以帮助我们理解周围世界的基础结构和原理，因此我们必须好好掌握这门学科。但是，遗憾的是，很多人发现数学非常难学，甚至觉得学数学是一件很痛苦的事，因此千方百计地逃避这门学科。2012年，英国有一篇关于世界数学日的文章指出：

> 我们还知道，很多成年人就是不喜欢数学，也不知道为什么要学习数学。在说到自己“数学学得不好”时，很多人一脸坦然，毫无羞愧之意。这种状况在世界其他地区并不多见。英国人对待数学的消极态度从孩童时期便开始了。有人认为，很多孩子在7—9岁时会遭遇对数学兴趣减退、成绩下降的问题，而且绝大多数孩子从此一蹶不振。他们的兴趣会发生转移，觉得数学枯燥乏味，到最后，他们会彻底放弃数学……随后，这个过程还会继续发生，兴趣不足与信心缺乏的问题又由父母传递给下一代。（斗胆提出一个问题：是不是也有老师参与其中呢？）

文章指出，英国人对待数学的态度尤其消极。而我认为，不仅在美国，在全世界很多地区都能看到同样的问题。其实，厌恶数学的态度古已有之。415年，奥古斯丁[注]说："我们面临的危险……是数学家与魔鬼订立了契约，他们要玷污人类的灵魂，把人类牢牢地羁押在地狱之中。"显然，他在几何课上也没有找到多少乐趣。（这句话有一定的误导性。通常而言，奥古斯丁对学习数学是持支持态度的。引言中的"数学家"一词其实应该译为"占星家"，但是，从中仍然可以看出很多人对数学持有消极态度。）

然而，如果教学方法得当，数学不仅学起来其乐无穷，而且在应用时可以发挥巨大的作用。数学的乐趣来源于数学难题和各种娱乐活动。当你的大脑为解决数学难题而飞速运转时（比如说，数学告诉你无穷大也有大有小），你就能深刻体会到数学的乐趣了。

在处理日常生活中的简单事务时，我们也许不需要掌握很多数学知识，而绝大多数人除了掌握一点儿算术知识以外，就对数学一无所知了。但是，在科学家和工程师努力了解事物的作用原理，并在此基础上构建产品时，数学就成了一个功能强大的工具，可以为他们提供灵感。没有数学，人们就难以理解自然世界并做出各种预测；没有数学，我用来写作本书的计算机以及为我们现代生活提供便利的其他科技也将不复存在。

首先，数学与自然界的运行规律密切相关。例如，数字与可触知对象可以相互匹配。但是，随着时间的推移，数字与现实逐渐分离开来。文艺复兴时期，数学家逐渐意识到他们是在玩一个规模巨大的游戏，他们可以自行制定规则，然后推动游戏的进程，并观察最终的结果。从此以后，理论数学就如脱缰野马般奔腾而去，只剩下应用数学与现实世界继续保持着密切的联系。数学家的想法与他们创建的数学世界，有时具有实际应用价值，有时则与实际生活没有任何联系。这两个不同结果的出现具有随机性（现在，这种情况也基本没有发生很

大的变化）。这个大型游戏有一个奇怪的特点：一方面，它是完全开放的；另一方面，它又表现出极强的限制性。数学研究覆盖哪些范围、可以制定哪些规则，全部取决于你，但是，一旦这些规则被认可，你就必须遵守。数学的世界里没有欺骗。

当我们考虑数字和现实的本质时，如果思维过于刻板，数学隐藏在深层次的随机性就有可能导致问题的发生。2015年，英国上诉法院（该法院的等级在英国法院系统中高居第二位）的三名法官在审理一个案件时，需要判断数字“1”的确切含义。当然，他们的判断与我们大多数人（甚至大多数的数学专业人士）的理解是不一样的。

这个案件是两家制药公司因为某种可以降低伤口敷料感染风险的化学溶液产生了纠纷，但是，令人意想不到的是，判决结果却导致英国法律修改了对数字“1”的定义。康维德公司生产的一种“浓度为1%—25%”的银基溶液受到专利保护，而它的竞争对手施乐辉公司生产的竞争产品是浓度为0.77%的银基溶液。后者认为，自己的产品不在竞争对手的专利保护范围之内。

早在2013年，这起纠纷就已经进入了审判程序。当时的法院认为，康维德专利保护范围下限中的“1”并不是表示“1”这个数值（即一个对象）。他们采用了一种在化学界比较常见但是在数学界并不多见的方法，认为这个值表示的是由两个有效数字界定的范围。根据这种“有效数字”规则，该一审法院认为0.95—1.5之间的所有数值都可以表示成“1”，而这种简单不对称的定义表明，施乐辉的产品没有触犯法律。上诉法院对这个说法感到不满，他们倾向于更为常见的就近取整法，也就是说，他们认为“1”代表0.5—1.499 9之间的任意数值。这个判决结果让施乐辉陷入了困境，也说明数学上的决定具有随机性。

除非我们坚定地认为1不多不少正好是1，否则“1”就会有不同的定义。因此，在律师看来，“1—25”的范围中包括0.5。上诉法院法官克里斯托弗·克拉克说：“语言学家可以认为‘一’表示‘正好是一，既不能

多，也不能少’。但是，对于经常从事艺术活动的人而言，这个字在不同的语境里可以突破整数的限制，表示一个数值范围。”他的这番话并没有多少意义，因为我们都不清楚他到底是怎么想的。

长期以来，数学家在处理数学问题时一直表现出明显的创造性（在这一点上，他们与律师截然不同）。有的公司为了开发新产品，允许员工尝试各种各样的想法和技术。在大多数情况下，他们不会生产出与商业世界有关的产品，但是他们偶尔也会推陈出新，设计出完美的新产品。数学家的研究与之有几分相似。比如，在他们开始研究负数的平方根（参见第8章），也就是虚数时，他们实际上是将数学这个游戏推向了新的发展路径。但是，随着事情的发展，他们为这些神奇的数字制定的各种规则，却在物理学和工程技术等领域发挥了巨大的作用。

在引入虚数的概念之前，没有哪一位科学家或者工程师能够做出这样的预言：“我们希望求得负数的平方根，因为它可以帮助解决我们所面临的问题。”同样，在设想出虚数之前，数学界也没有人会这样想：“我们怎么解决摆在物理学家面前的这个问题呢？”数学家在提出新的概念以及制定一套相应的规则之后，通常不会很认真地考虑他的行为会引发哪些后果。但是，一段时间之后，各种应用就会应运而生。

总的来说，在19世纪之前，只要不是严重缺乏数学天赋，任何人都可以比较熟练地掌握数学这门基础的自然科学。根据我的经验，只要能学会分数的运算（很多人怎么都学不会，而且人数之多，实在令人吃惊），你就可以学会包括微积分基础（这门课程从名称上看似乎挺简单，但其实难度不小）在内的所有内容。但是，到了19世纪，数学领域发生了两个变化，从此以后，普通大众与科学之间就多了一道鸿沟。

第一个变化是，人们在数学中使用的方法越来越复杂，学生必须在课后投入大量的时间，才能掌握这些方法。比如，你可以随便挑选一篇现代物理方面的论文，文章中可能至少会有一个在中学数学中学不到的数学方法。爱因斯坦在研究广义相对论时，需要别人帮助他解决数学上的问题。这件事一点儿都不奇怪，因为数学本身非常难，超出了爱因斯坦的能力范围。他是自然科学领域的大师，但数学却是他比较陌生的领域。

让数学变得曲高和寡的第二个变化是它与科学研究的关系发生了逆转。在19世纪之前，数学一直是为科学研究提供服务的，但是到了20世纪，数学逐渐占据了主导地位。例如，人们从数学中的对称性出发，试图将自然界中各种作用力的研究统一起来，但是，随着研究的深入，人们越来越难以摆脱对技术术语的依赖，从而将外行人拒之门外。矩阵力学也是一个类似的例子。人们利用矩阵力学来解释量子行为，但是，这种方法很抽象，所以对于当时的科学界而言，矩阵力学基本上是一个非常陌生的内容，更不要说普通人了。最终，留在他们记忆中的就只是一堆数字和运算法则。

数学上取得的发展本身没有任何问题，但不幸的是，它们背上了沉重的包袱。如果普通人无法清楚地了解你所研究的科学，他们就不会轻易地同意政府将纳税人的钱投入这个领域。20世纪90年代，美国政府需要在资助国际空间站（ISS）还是支持超导超级对撞机（SSC）这两个方案中做出选择，他们做出的最终决定遭到了众多物理学家的诟病。SSC是一种大型粒子加速器，当时，已经在得克萨斯州启动了建造工作。诺贝尔奖得主、物理学家史蒂文·温伯格（Steven Weinberg）指出，SSC的规模比欧洲核子研究中心（CERN）的大型强子对撞机（LHC）还要大，研究水平领先后者有10年之多，研究成果必将有助于进一步了解宇宙的基本规律。最终却是ISS得到了资金支持，而SSC项目则被迫取消。与SSC不同的是，ISS可以帮助我们进一

步探索太空和太空旅行的秘密，也许在未来会非常有价值，但是对于科学研究几乎没有任何贡献。

而且，截至目前，美国政府投入到ISS项目上的经费是SSC项目的10倍，而进展却十分缓慢。但是，政客们都非常清楚ISS项目到底是干什么的，负责拨付资金的人也可以听懂这些内容。凭借不同于SSC项目的这个特点，ISS项目在这次竞争中胜出。对撞机项目太复杂了，让那些政客理解相关内容太困难了。因此，科研资金支持的是对科学研究没有任何益处的ISS项目，而不是有助于美国在物理学领域保持世界领先地位的SSC项目。

最后，CERN和LHC赢得了本来属于SSC的荣誉。然而，从一定程度上看，这不过是一个毫无意义的胜利，因为他们同样面临那个难题：如何解释清楚LHC到底是干什么的。所谓的“重现大爆炸前的环境”、“寻找希格斯玻色子”，到底是什么意思呢？他们在使用这些术语时已驾轻就熟，但是向普通大众解释它们的确切含义并不是一件易事。通常，人们将希格斯玻色子描述成“使其他所有粒子具有质量的粒子”。从某些方面看，这个说法不能算错，但是它也不全对。如果不借助数学方法，我们甚至难以说清楚这个说法到底错在什么地方。问题是，一旦使用了数学术语，听众就会兴趣全无，再也不想听下去了。

我希望读完本书之后，大家不会产生数学难学的消极思想。本书将探讨数字是否真实存在的问题，在这个过程中，大家会发现，数学不仅饶有趣味，而且非常有用。数学在现代社会、现代科技的发展过程中起到了核心作用，但是，我们不应该因此对那些基础性问题视而不见。现代科学是不是过于强调数学的重要性了呢？马克斯·泰格马克等科学家甚至认为我们所在的世界就是一个数学世界，数字不仅真实存在，而且是促使世界运行的根本原因。这些科学家是否将数学与现实混为一谈了呢？数学到底是宇宙的核心内容，还是一个可以帮助我

们了解周围世界的有效工具呢？在探讨数学是如何取得科学世界霸主地位的同时，我们就能找到这些问题的答案了。

但是，我们首先必须回到数学思维的起点。我们将沿用数学书籍的惯例，先做一个规则自定的游戏：假设我们可以发明数字。

---

1. 奥古斯丁（354—430），古罗马基督教思想家，教父哲学的重要代表人物，著有《忏悔录》《论自由意志》《论三位一体》等。——译者注

第2章
史前人类的计数系统

Are Numbers Real?

你觉得有没有必要数一数自己的孩子，以便确定他们是真实存在的？应该没有必要吧。同样，人类社会早期的采猎者没有大型项目或商业活动的概念，对于他们来说，数字几乎没有任何意义。但是，随着人类定居下来并开始从事贸易活动，计数和记录结果的能力变得重要起来。首先，我们可以采用阿尔伯特·爱因斯坦的方式，通过思想实验的形式来理解这个问题。假设数字还没有出现，而我们的任务是发明这些数字。我们并不确定历史上数字是如何被创造出来的，但可以推测出这个历史过程的大概情况。

我们从计数开始。我们生活在一个高度重视数字的世界之中，因此，如果有人告诉你，计数其实并不一定需要有数字，你也许会觉得这个说法荒谬可笑。但是，事实确实如此。在研究集合论和无穷大时，我们就会遇到这种情况，因为我们经常会见到可数无穷大和不可数无穷大这样的概念，尽管两者都不指数字。我们暂且不考虑这些复杂的概念，而是集中精力了解做记号的计数方式为什么不需要借助数字。

假设我是一名史前农民，我生活的社会没有数字。邻居向我借山羊，我答应了他。（我不知道邻居为什么要借山羊，我对山羊以及史前农民也不甚了解。）我和邻居是朋友关系，我很信任他，但在朋友归还山羊时，我仍然希望找到一个办法，可以确定他如数将山羊还给了我。因此，我把手掌张开，五指伸直。在邻居从我的羊圈里赶出第一头羊时，我把小拇指收回到手心的位置。（小拇指弯曲时，无名指往往会随着小拇指一起弯曲，因此你可能需要用另一只手协助小拇指。掰手指是一种比较低级的计数方式，但是十分方便。）第二头羊离开羊圈时，我收回无名指。就这样，当第五头羊被赶出羊圈时，我把大拇指收回来，横扣在其他手指上面。这时候，邻居觉得羊已经够了。

几周之后，他来还山羊。当这些山羊被赶入羊圈时，我通过同样的方法，统计了山羊的数量。最终的结果一样，我知道借出去的山羊已悉数归还。（严格地说，我不知道还回来的这些山羊是不是我借出去的那些，但在这里我们不考虑这个问题。）当时，我不知道我借出了多少头山羊（我没有“多少”的概念，也没有数字的概念），但是我知道邻居没有欺骗我。事实证明，计数是一种非常有用的工具，可以帮助我们解决身边的问题。

有证据表明，古时候，人们在采用这种计数方法时使用的工具是符木。已知最早的符木是一根有刻痕的骨头，称作伊尚戈骨，可以追溯至20 000年前。这块狒狒的腓骨（小胫骨）上刻有三组深深的刻痕，每组的和分别是60、48和60。这些符木很可能就是计数用的。它们不仅可以表示更大的数，而且计数结果可以长时间保存，因此是一种优于掰手指的计数工具。

我们只能推测伊尚戈骨是用来计数的，但是无法找到背景资料加以确认。那些刻痕也可能是一种装饰。但我们可以确定，在距今更近的史前时代，人们使用了大量的符木标记，而且这些标记显然是用于

记录的。我们可能永远无法确定符木这种无数字计数工具第一次出现的时间，但我们知道把符木作为一种常用的计数工具已经有漫长的历史了。然而，我们设想的发明数字的实验才刚刚开始，它远不像制作一根符木那样简单。

接下来，我们设想，在把山羊借给邻居后的第二天，我女儿问我哪些山羊被借走了。但我已不记得朋友借的是哪些山羊（一段时间之后，所有山羊在我的脑海里都变成一模一样的了）。于是，我一面说“就是……”，一面掰起手指头。这已经是我能找到的最有效的记录方式了。若干天之后（在这期间，邻居又找我借了几次山羊，看来他有向邻居借东西的习惯），我突然灵光一现。每次说到借给邻居的那些山羊时，我为什么都要重复掰手指这个无聊的动作呢？只要说“一只手的山羊”，不就可以了吗？如果他需要多借一头山羊，那我怎么表示呢？“一只手加一根手指的山羊”。于是，我不知不觉就发明了数字。由于这些数字都是根据我的手指发明的，因此它们可以叫作“手指数”。

不幸的是，史前版的我有点儿上年纪了，记忆力也衰退了。在一段时期里，由于邻居借羊的频率异常高，所以我需要通过在符木上刻痕的方式，帮助自己记住借出去了多少头羊。然而，我发现，既然“手”这个词可以用来表示一定数量的手指或者符木刻痕，那么，如果我用某个特定的刻痕或者图案来表示手，计数结果不就一目了然了吗？那样的话，我就不需要将很多刻痕转化成若干只手了。刚开始的时候，我把符木刻痕画成手指的样子：



但是，过了一阵子之后，因为懒惰，我把这个刻痕做了简化处理，用一条横线表示大拇指，用一条竖线表示其他手指，于是这个图

案就变成了一个不规则图形：



在书写和惰性的双重作用下，这些数字逐渐形成了一套独有的符号体系。这套数字系统简单易懂，即使你之前从未使用过，也从未见过，看到下面的符号之后，你也应该能立刻说出这个数是多少：



没错，这个数就是现代数字系统中的12。然而，现在讨论这个数还为时过早。在史前人眼中，这个数就是“手—手—手指—手指”。但也有可能是“手指—手指—手，手指—手指”，这是因为我可以利用左手的手指数出这个数包含多少只手，还可以利用右手数出余下的手指数。太棒了！我是不是已经成为一名数学高手了？暂且还算不上。但是，我已经是一名算术师了，如果真有“算术师”这个词的话。（但好像真的有这个词，因为电脑的拼写检查程序没有报错。）的确，这些内容太简单了，可能还不足以称为数学吧。但是，在讨论更复杂的情况之前，我们先看一看，到现在为止，我们到底掌握了哪些技能。

我发明的这种符号系统是用数字表示实物，也就是说，这些数字是对现实世界中具体事物的直观表示。具体地说，在本例中，这些数字表示的是山羊的数量。对于现代人而言，既然这些直观的符号可以表示山羊，就一定可以表示玉米等其他事物。但是，我们知道，早期的准数学家在历经了一番周折之后，才艰难地完成了这个由具体到抽象的飞跃过程。事实上，数字的通用性（指数字与实物剥离，变成独立的符号。正因为数字的这个特性，我们现在不仅可以用“4”表示香肠，还可以用它表示汽车）并不是与生俱来的。

就像数山羊这个思想实验一样，古时的乌鲁克城邦居民在学会书写之后，也经常需要考虑计数的问题。乌鲁克是最早的城市之一，在伊拉克还能看到该城遗址。公元前4000年，乌鲁克在苏美尔文明中占据核心地位，存世2 000多年。但是，乌鲁克的居民没有发明出可以表示所有事物的数字系统。他们虽然进行了一定程度的归纳，却认为有的事物与其他事物相差甚远，在表示这些事物时需要使用特殊的数字。例如，他们使用一套数字系统表示人、活的动物和干鱼（不要问我为什么），同时使用另一套数字系统表示谷物、奶酪和鲜鱼。

但是，在应用过程中，有人不可避免地会在某些场合中突破限制，用某个事物（比如山羊）独有的计量系统表示其他事物。于是，数字逐渐具有了通用性。我之所以用大量篇幅讨论这个过程，是因为它对于回答"数字是一种真实的存在吗"这个问题具有非常重要的意义。如果事实证明数字并非真实存在，那么为什么它们可以如此好地表示现实呢？对于这个问题，美国数学家理查德·汉明说：

> 抽象的整数不仅可以用于计数，而且效果非常好，这实在令我吃惊。我曾经试图向朋友们介绍我的这种心情，但是他们几乎都无法理解。6头绵羊加上7头绵羊，就有13头绵羊，6块石头加上7块石头，就有13块石头，这样的结果难道不令人吃惊吗？宇宙间竟然有像数字这样简单的抽象概念，难道不是奇迹吗？我认为，这个事实强有力地证明了数学的神奇性达到了我们难以想象的程度。我认为数学既不可思议，又难以解释。

接着说我们的思想实验。过了一段时间，我可能会调整掰手指这种计数系统，并且使用一些有形的象征物。这些有形的象征物可能是计数石，也就是"calculi"［小卵石，"calculation"（计算）与"calculus"（微积分）即由此演变而来］，也可能是算盘珠，还可能是我们今天仍在使用的硬币。事实上，在发明数字之后不久，我肯定就需要制作出某种有形象征物。从记账这个角度看，我发明的这种书

写符号没有任何问题。例如，看了这些符号，我就知道我借给邻居多少头山羊。这种记账方式是可行的，因为邻居和我是朋友，我们彼此信任。但是，如果我是一个不诚信的人，我就可以在符木上添加两条刻痕，而且看不出任何破绽。

邻居来还山羊时，等到数完一只手的山羊之后，我会装出一副很伤心的样子说："你只还给我一只手的山羊，还有'手指—手指'的山羊没还呢？"同时，我会把修改过的符木拿给他看，脸上则是无辜又可怜的表情。邻居不能复制我的符木，因此他可能没有办法为自己辩解，他要么多还给我两头山羊，要么把我痛揍一顿。

实际上，到目前为止，用"手指—手指"来表示"2"的计数方式可能已经让我不胜其烦了。因此，富有独创性的我可能会想出一些字词，用来表示介于手指与手之间的数值。经常与文字打交道的人都希望字词简短易记，同样，我也希望这些字词不要太长，所以我最后想出来的字可能是："芬，戈，纽，喀，手"。于是，我一面适度地做出伤心的表情，一面质疑邻居："还差戈头山羊呢！"

无论我如何表示剩下的这两头山羊，不经意间，我毫不费力地完成了一种全新的数学运算。在我作弊之后，我的符木显示出山羊的总数是手—戈（实际上，我可能会把这两个字连起来，表示"7"这个数）。邻居还给我一只手的山羊，还欠我戈头山羊。从手—戈头山羊中收回一只手的山羊，还剩下戈头山羊。也就是说，手—戈减去手等于戈。这难道不是一道算术题吗?

因此，如果我不诚信，我就可以借助符木和我新发明的骗术进行骗人的勾当。幸好，聪明的邻居发现，我的符木仅仅通过刻痕表示山羊的数量，而这些刻痕改起来非常方便，因此他有被骗的危险。于是，他拿出了一套具体的象征物，每枚象征物代表一头山羊。这些象征物是他自己制作的实物，易于辨认，而且我难以复制。在他归还山羊时，我收到一头山羊，就还给他一枚象征物。归还了一只手的山羊

之后，这些象征物就全部回到了邻居手中，因此我无法欺骗他。但是，新的问题又出现了。由于邻居经常来借山羊，我有点儿不高兴了。因此，我们达成了一个新协议。在邻居再次借山羊时，作为补偿，我可以留下一枚象征物。将来，我可以利用这枚象征物从邻居那里换取某些物品，例如一袋面粉。

由于彼此之间的不信任，再加上非常简单的算术运算，钱在不知不觉之中便诞生了，并且为有偿服务的产生奠定了基础。由此可见，数字的功能强大到令人瞠目结舌的程度。但我们还是从金钱回到纯算术这个话题，继续考虑“手”的含义。刚开始的时候，我可能只会用这个概念来表示羊群的大小，但是，我很快就会发现它的通用性，也就是说，它还可以方便地表示苹果、人、鱼叉等事物的数量。作为一个数字，手可以用于很多方面，它可以告诉我们某种事物到底有多少，无论这种事物是什么。

到目前为止，这就是手的全部功能，而且这些功能已经足够满足我们计数的需要了。坦白地说，这种状况将维持相当长的时间。在清点存货、借钱、购物和销售等活动中，手可以发挥不可思议的显著作用。在准备晚餐时，手可以告诉我们有多少人将与我们共进晚餐。手还可以告诉我们，多少天之后，白天会慢慢变长。正因为如此，当简单的符木演变成书面数字之后，古巴比伦人发明了非常先进的六十进制，这是最早出现的计数系统之一。

“六十进制”是指在60这个位置，所有的数字都要进位到下一级。我们现在使用的十进制数字系统，很有可能源于双手一共有10根手指这个事实。（从技术上看，本章讨论的手数字系统是五进制。）古巴比伦人的这套数字系统是楔形文字，是用尖笔在陶片上刻画出来的。值得注意的是，这是一个出现时间非常早的位置计数系统（某个数字在一排数字中的位置不同，它所表示的值也不同，例如，“1”表示的可

能是“1”，也可能是“60”或者“3 600”），直到2 000多年后，位置计数系统才成为一种常见现象。

五进制、十进制和二十进制都比较好理解（考虑二十进制时，我们可以加上脚趾），但是六十进制乍一看似乎比较奇怪。其实，60这个数字使用起来十分灵活，它可以被1、2、3、4、5、6、10、12、15、20和30整除，因此在除法运算中非常方便。大家不要彻底否定六十进制。别忘了，我们在将秒换算成分钟、将分钟换算成小时以及表示角度时，仍然在使用它。古巴比伦人将楔形文字书写在陶片上，因而留下了大量数字，其中有很多是用来记账和管理交易的，但是还有一些数字则是古巴比伦人研究天空时留下来的。我们知道，古巴比伦人对天空进行了深入细致的研究，主要是由于占星术的发展。

我们回过头来接着讨论史前农民这个思想实验。我们创造的数字都是用来代表实物的，如果不用来表示一个真实的物体，这些数字就毫无意义。它们更像形容词，而不是名词。我不能给你手—戈，我只能给你手—戈头山羊或者手—戈只篮子。我也没有办法让你看到手—戈。也许你认为，我可以用合适的符号把手—戈这个数字表现出来。但是，就像山羊的画像不是山羊一样，表示手—戈的符号也不是手—戈这个数字。时至今日，相当多的人都把数字理解为形容词，其中有很多都是那些上学时看到数学就头疼不已的人。这是因为，直接表示具体物体只是数字最基本的功能。

很多早期数字系统使用起来非常不方便，远远落后于古巴比伦人的发明，原因就在于他们在数字与实物之间建立了这种联系。例如，古希腊人用标准字母表示数字，但是，他们必须重新启用一些早已废弃的字母（例如与大写的“F”非常相似的第6个古希腊字母）才够用。这种系统要求在使用时必须根据上下文区分字母与数字，因此会造成混淆，但这个事实说明，写作与记账这两种活动在人类社会早期是严

格分开的。这个结论似乎是在研究腓尼基人时得出的，在研究希伯来人早期的数字表示方法时得到了证实。

我们大多数人更为熟悉的罗马数字不仅使用了简单易懂的符木标记，还使用了与希腊数字相似的字母系统。实际上，罗马数字与我们在前文中想象的手进制数字系统有一个明显的相似之处，我们可以把罗马数字中的I和V分别视为表示手指和手的符木标记。希腊数字用不同的字母表示10和100的倍数，而罗马数字则直接采用字母重复出现的办法，同时，他们还通过有趣的位置变化，表示某种特殊意义。在罗马数字中，所有字母都是按照数值由大到小的次序排列的，因此，如果一个比较小的数字出现在比较大的数字之前，就表明这个小的数字应该被减掉，因为它仅仅起修正作用。

例如，大家想一想钟面上的罗马数字。在“4”这个位置上的数字是什么样子？熟悉罗马数字的人都知道，对应的数字是“IV”，其中“V”表示5，而“I”则表示“减去1”，因此整个数字表示4。经常有人错以为罗马数字中的4是IIII这种更为简单的表现形式，历史上也确实有过这个奇怪的习惯，在表盘和钟面上用IIII表示4（尽管9仍然被表示成IX）。但是，看到罗马数字时，我们的理解通常都是正确的。

在我们看来，古罗马人和古希腊人使用的这两套系统极不方便。与古巴比伦人相比，他们确实落后了。的确，人类大脑在处理六十进制的数字时有些麻烦，这个事实也反映了人类短时记忆的特点——人脑一次只能处理七八条信息。正因如此，人们习惯在电话号码（通常超过7位数）中间插入空格。相较于六十进制，五进制和十进制都更容易处理。但是，出现时间更早的六十进制在很多方面具有后来者无法比拟的优势。

当然，罗马数字几乎不占据任何优势，因为它们实在太笨重了。大家可以比较罗马数字中的年份，如1999年对应的罗马数字是MCMXCIX。（罗马数字偶尔也会占上风，例如2000年的表示方式较

为简洁——MM。）大家也许会感到奇怪，我们为什么偶尔还会使用罗马数字呢？我认为，这可能是因为截至20世纪，人们一直对古典文化存有莫名其妙的敬畏之心。一些多年来一直被视为丑陋不堪的古典建筑风格重新流行起来，原因也在于此。与阿拉伯（印度）数字相比，罗马数字的唯一优势就是曲线比较少，刻在石头上比较容易。

更令人意想不到的是，罗马数字还有一个非常严重的问题：它们会大大增加算术基本运算的难度。比如，若计算XXIII和XLV的和，使用罗马数字你将无法找出一条简单法则（数学界青睐的就是简单法则）。但我们可以很容易地教会孩子们如何求23与45的和，这是因为我们使用的是一个位置计数系统，从数字所在的列就可以看出它是10的多少次幂。根据这个特点，人们总结出了算术的运算法则（参阅第6章）。罗马人在科学上几乎毫无建树，与他们使用不方便的罗马数字之间可能脱不了干系。

就像邻居借羊时使用的最原始的符木一样，古时的希腊数字和罗马数字都不是数学工具，人们也不会像后世的数学家那样，利用这些数字从事数学研究。尽管古希腊人有研究数学和科学的传统（我们很快就会讨论这个问题），但是他们对数字的理解不同于现代人。我们暂且不讨论这个问题，而是继续我们的思想实验，看看史前的那位山羊主人可以利用符木系统完成哪些基本活动。

你也许认为负数对于这些早期算术师而言还是一个非常遥远的事物。按照现代人的理解，事实确实如此。你不可能有–2头山羊，我也拿不出–2个物品，无论这个物品是什么。因此，–2不可能直接表示现实世界中的任何对象。然而，在记账时，–2却是一个重要的概念，即使这个概念在刚开始的时候没有表示成负数的形式。作为一名史前农民，我把一只手的山羊借给邻居，那么在邻居归还之前，我的羊群就少了一只手的山羊。尽管我记录的是缺少一只标准手（正值）的山羊，但是实际上，符木上的刻痕或者弯曲的手指都表示山羊的数量是

负数。邻居归还山羊时，每归还一头（正值）山羊，我就会填平符木上的一条刻痕，直到5条刻痕都被填平。第一次使用符木时，我们在不使用数字的条件下学会了计数，而现在，在没有意识到负数这个概念的条件下，我们学会了利用符木进行负数的运算。

即使山羊被借走了，它们依然是整个事件的核心。无论是负值还是正值的手，表示的都是位于现实世界某个地方的一群实实在在的对象。但是，数学家若要自由翱翔，就必须切断与世俗之间的联系。

不同的文明意识到这个问题的时间有先有后，但是在西方传统中，古希腊人是第一个觉醒的。由于希腊人没有采用世人一致认可的科研方法，而且犯了很多错误，所以现在很多人在提到希腊的科学水平时都多少有点儿蔑视之意。希腊人在数学界的影响力也十分有限。但是，即使他们毫无建树，我们也应该感谢这个民族，因为是古希腊人促使数学取得了有史以来最重要的一个进步。他们清楚无误地宣告，数字不一定非得与现实世界中的具体对象相对应。某个古希腊教派深信，整数突破了物品数量统计的羁绊，具有一种全新的意义。

第3章
毕达哥拉斯：万物皆是数字

Are Numbers Real?

传说中，毕达哥拉斯学派将一条座右铭刻在门楣上：万物皆是数字。毫无疑问，我们大多数人对毕达哥拉斯学派都非常陌生。关于这个古希腊学派，我们唯一确定的可能就是毕达哥拉斯定理，即直角三角形斜边边长的平方等于其他两边边长的平方和。其实，这条定理的基本概念早在毕达哥拉斯之前就已被人提出来了，但是，定理的证明可能要归功于毕达哥拉斯或者他的门徒。

几何学（我将在第5章深入讨论）似乎是人类继整数计数之后最早探讨的一个数学领域。我们不清楚几何学起源的确切时间和具体过程。大多数古希腊人认为几何学源自埃及，但其实一些描述简单几何关系（如毕达哥拉斯定理）的相似概念，在巴比伦、美索不达米亚和东方出现的时间更早。古埃及把从事几何研究的人称作“司绳”（rope stretcher），暗示这门学科与从事建筑测量及土地分割的人员有关。几何学之所以成为人类较早研究的学科之一，从这个表达上也许可以略知一二。［现代英语中的“几何学”（geometry）一词源于希腊语中“地球”和“测量”这两个词。］但是，毕达哥拉斯及其门徒研究的并不是这种需要亲力亲为的实用数学。

毕达哥拉斯出生于公元前570年左右，古希腊的数学研究就是从那个时期开始的。据说，毕达哥拉斯去过埃及。他从埃及人那里汲取了一些想法和概念，后来都被纳入毕达哥拉斯学派的信仰体系，该学派的主要成员被人称作数学家（mathematikoi）。前文的思想实验告诉我们，数字的起源可能与现实世界密切相关。毕达哥拉斯学派的功劳则是将这个概念提升到基本法则的高度，但与此同时，他们也将数学与简单数字区别开，使数学有可能摆脱束缚，无须直接应用于我们周围的物质世界。

毫无疑问，毕达哥拉斯学派超越了计数的限制。他们认识到图形的重要作用，并将图形变成一个重要工具，从而拉开了现代数学与现代科学的漫长历史的序幕。从一定程度上看，所有人都会总结规律（事实上，对周围世界有反应的所有生物都会如此）。如果做任何事都需要从头学起，那么人生的复杂程度就会远胜当前。但实际情况并非如此。借助规律，我们不仅可以认识周围的世界，还可以对探知的一切做出反应。例如，假设我设计了一个可以打开卧室窗户的机器人。事实上，所谓的卧室窗户其实是一扇门，正对着一个法式小阳台。因此，我设计的这个机器人必须学会把钥匙插进位于墙上某个位置的锁孔，然后转动钥匙，再按下位于另一个位置的门把手，最后推开门。不仅如此，它还必须控制好力量的大小，以免损坏这扇门。

我家客厅的窗户与之相同，但是位置不一样。如果我把这个机器人搬到客厅，那么尽管我编写的程序十分精确，但它仍然会找不到目标，也无法完成打开窗户这个任务。当然，如果客厅采用的是上下推拉式窗户，那么后果就会更加严重。在掌握门窗等事物的规律之后，我们就可以有针对性地采取应对措施，而不需要学习每一扇窗户的打开方法。

规律在我们人类了解宇宙奥秘的过程中发挥了同样的作用，所有科学都在利用规律的基础上简化了理解过程。如果必须逐个地研究所

有原子，我们将永远无法探知宇宙的秘密。但是，如果我们可以找出原子的一般规律，并将这条规律应用于所有原子上，我们的研究就可以取得进展。

在毕达哥拉斯学派的鼎盛时期，古希腊人还没有提出原子的概念（“原子”一词源于希腊语中的atomos，意思是“不可分割的东西”），但是他们仍然强烈地感受到了规律的重要性，包括几何图形的规律、音乐和弦的规律和数字的某些规律。然而，他们对规律的利用达到了过犹不及的地步，这也是人类经常犯的一个错误。我们非常善于无中生有，总结出某种根本不存在的规律，包括视觉规律（例如，从一片阴影中看出一个妖怪）和统计规律（例如，我们总希望从一系列事件中总结出前因后果，即使这些事件纯粹是随机事件，毫无联系可言）。

别忘了，毕达哥拉斯的这些门徒不仅仅是数学家，还是一个学派的成员，有独特而奇怪的信仰，例如，他们特别反感吃豆子。据说，这种信仰源于某些神秘的象征意义，因为豆子的外形与人类的某个器官相似，所以对于他们来说，吃豆子的行为与同类相食没有多大区别，这种行为对于这个素食主义教派而言是不合适的。

在毕达哥拉斯的门徒看来，很多重要的规律都与整数密切相关。他们认为，宇宙万物的基础是数字，数字不仅是人类的发明创造，还揭示了现实世界的基本架构。与所有凭借一己之力在这个世界中谋取立足之地的生物一样，数字也被他们赋予了各种属性。例如，数字1与思想及其独一无二的特性相关；数字2表示意见，因为他们认为意见是需要分享的；数字3与完整性有关，这是因为所有完整的事物都必须包括开端、中间和结尾，具体的事物需要三个维度才能定义它们的物理存在。

就这样，他们发现了一条又一条规律。例如，由于奇数被视为阳性的，而偶数是阴性的，所以数字5表示婚姻（这里的“婚姻”一词可能

是维多利亚时期的委婉语），因为它把第一个真正的奇数3（他们认为数字1过于独特，因此不应该归属于奇数的行列）和第一个偶数2融为一体。对数字的这种联想在数字10上达到了巅峰，他们认为10表示完美。10不仅是1、2、3、4这4个数字之和，而且10个点可以排列成一个完美的等边三角形。

毕达哥拉斯学派还从音乐中找出了数字规律。他们不仅发现音乐节奏有某种规律，还发现乐器的琴弦或管腔长度必须符合某种规律，才能演奏出悦耳动听的乐曲。例如，把琴弦加长一倍，演奏出的音调就会降八度。此外，他们还找到了演奏其他悦耳和声所需要的琴弦比。这种以数学方式研究音乐的行为似乎无足轻重，但是它一直被视为人类第一次利用数字得出某种科学法则，因此成为科学发展史上的一个重要里程碑。在此之前，人类对自然的研究全部是定性研究，而从此以后，定量研究也登上了历史的舞台。

过分相信数学与现实之间的联系，并以此为依据进行演绎推理，是一种非常危险的行为。（至少，亚里士多德是这样认为的，这位做不到绝对公允的哲学家也对这个古老的学派提出了严厉的批评。）一位名叫菲洛劳斯的毕达哥拉斯门徒就犯了这个错误。毕达哥拉斯学派对数字10的完美性倍加推崇，对此深信不疑的菲洛劳斯因此认为，宇宙间必然还存在一颗不为人知的行星。他把太阳、月亮、地球、已知行星以及恒星的数量相加，即1 + 1 + 1 + 5 + 1，得到的结果是9。据说，菲洛劳斯认为，由于宇宙必须处于完美平衡的状态（只有这样，才与毕达哥拉斯学派深信不疑的规律相一致），所以所有这些天体的数量和必须是10，这是由数字10的重要性决定的。他因此推断必然存在一颗未知的行星。

现在，人们很容易将他推断存在的这颗行星与当时尚未被人类发现的天王星（我们就不提海王星了）联系起来，但是他提出的另外一个观点却更加疯狂（从物理学角度看，也更加不可能）。他认为，宇

宙中还有一个与地球相对应的“反地球”。当时，我们更加熟悉的希腊宇宙论（在这套理论中，宇宙与太阳系相似，但地球是宇宙的中心）还没有出现，关于宇宙结构的普遍观点是宇宙中心有一团火，构成了整个宇宙的光源。他们猜测，反地球就在这团火的另一侧，因此我们永远看不到它。

尽管科幻小说中经常会重现这样的行星［注意，不要与祝融星（Vulcan）混淆。人们曾经假想在水星的轨道内侧还存在一颗行星，即祝融星］，但是天文学从来没有找到它确实存在的证据，它只是人们根据宇宙运行规律的数学模型进行演绎推理的产物。尽管现代物理学也会进行这样的演绎推理，但是我们使用的数学方法已经得到了不断完善，而且在提出某个假设之后，我们还会通过观察或者实验，验证这种假设是否成立。然而，在菲洛劳斯那个时代，人们不可能利用这个方法验证反地球理论的真伪。

尽管也遇到过一些小的挫折，但是规律在人类探索宇宙本质的活动中占据了绝对优势，所以这种演绎推理的方法必然会受到重视。有的规律非常明显，你是不可能视而不见的。比如，我们都非常熟悉昼夜交替的规律。在时间方面，有些规律［例如一天可以分成24个小时，一个星期有7天（这些规律都是基于太阳、月亮等星体的运转形成的）］是人类总结出来的，仅在人类使用这些规律时才有意义，而有些规律（例如天、年等）则是自然现象，是真实存在的自然规律。

太阳的运转遵循某些规律（运动方向始终不变，速度大致相同）。在更大的时间跨度里，太阳的高度变化，以及行星和恒星的光芒在天空中的重复性变化，都会表现出一些规律。此外，季节性交替也会表现出一定的规律。在时间跨度更大的情况下，生命本身也会表现出某些规律。因此，古希腊人借助各种规律来理解周围的世界，也在情理之中。

所谓的亲和数[注]，例如220和284，可能是源于毕达哥拉斯学派的另一个概念，虽然它源于数字之间的浪漫关系，但其实并没有那么夸张。这两个数字彼此之间“含情脉脉”（毕达哥拉斯学派认为这两个数字之间充满了爱情的味道），甚至引起了珠宝商的兴趣：他们设计的心形金属首饰被分成两半，上面分别刻有数字220和284。

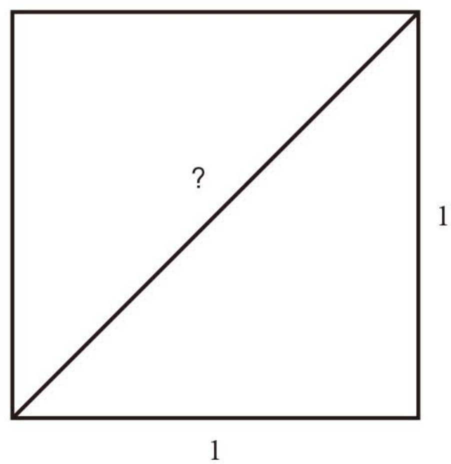
对亲和数的探索是一场冒险之旅，有可能把你变成资深数学呆瓜。列出220的所有因数（即可以整除220的数），也就是1、2、4、5、10、11、20、22、44、55、110、220。不算220，把剩下的数相加，和为284。与220相比，284的因数则要少得多，仅有1、2、4、71、142、284。不算284，剩下的数字之和是220。在痴迷于整数的毕达哥拉斯学派看来，这种关系具有让人无法抗拒的魔力。

在我们看来，这种关系仅是一个有意思的现象，在人们问起“为什么要用220这个数”时你可以卖弄一番，也可以把它变成一个私人的小秘密。但是，在把数字视为万物基础的毕达哥拉斯学派看来，这些数字显得尤为重要，应该被供奉在重要数字[注]的神殿之中。在探索宇宙奥秘时，如果完全依赖于数学（具体地说，就是整数），就会眼前漆黑一片，看不见前方的道路。

很显然，这就是数学伊甸园里的一条蛇。据传，有一个人不小心闯进了这个整数统治一切的王国，结果丢掉了性命。

不难想象，深信数字具有重要意义的毕达哥拉斯门徒会不惜一切来维护他们的信仰。传说中，他们确实也是这样做的。受害者名叫希帕索斯，是毕达哥拉斯学派的成员之一，因为胆敢将他们发现的“肮脏”的数学秘密公之于众而遭到杀害。可以想象，如果某个主流教派突然发现，他们的某一段经文宣称该宗教建立在子虚乌有的基础之上，那么他们肯定会想方设法加以掩饰的。而且，整个教派肯定会坐立不安。

一切都源于毕达哥拉斯提出的那条显然正确的定理。当人们利用这条定理计算单位边长正方形的对角线长度时，问题出现了。这个正方形的边长可以是实践活动中的任何单位长度，可以是1厘米，也可以是1英里(注)。但是，方便起见，我们可以认为它的长度就是1。当然，这个长度可以是任意值，然后我们可以将这个长度定义成一个新的长度单位，从而把正方形的边长变成单位长度。



我们画出了这个正方形的对角线。到目前为止，没有任何问题。根据毕达哥拉斯学派最喜欢的那条定理，可以很容易地计算出这条对角线的长度，因为它是直角三角形的最长边（事实上，我们可以从两个全等的直角三角形中任选一个）。也就是说，借助数学这个功能强大的工具，我们求出直角三角形另外两边的平方和，就可以得出这条对角线长度的平方。由于我们将正方形的边长定义为1个单位长度，因此对角线的平方就等于$1^2 + 1^2$，也就是$1 + 1$，结果等于2。同样，我们可以方便地把这条对角线的长度（也就是我们试图计算的值）定义为$\sqrt{2}$，即2的平方根。这个数的平方是2。

整个过程似乎没有任何问题。但是，当毕达哥拉斯学派的门徒试图计算$\sqrt{2}$的确切值时，情况一下子变得复杂起来。别忘了，在他们的心目中，整个宇宙的基础就是整数。因此，所有的数字，包括$\sqrt{2}$，都应该可以通过整数表现出来。$\sqrt{2}$显然不等于1，因为1的平方等

于1，而不是2。同样，$\sqrt{2}$也不可能等于2，因为2的平方是4。这不成问题，因为毕达哥拉斯学派的门徒知道在1和2之间还有其他的数，这些数是两个整数的比，也就是我们现在所说的分数。

如果古希腊人真正理解了分数的概念，毕达哥拉斯学派在数字抽象化进程上的步伐就会迈得更大一些，而不会将整数的存在与现实世界中具体事物的数量统计割裂开来。在我们使用正整数时，这些数字可以随时回归到具体事物的数量上来。比如，如果我有3头山羊，你牵走1头，那么我只剩下2头山羊。但是，如果朝着简单分数（例如一半，即1/2）迈进一步，这些数字就再也不能同样完美地表示现实世界中的事物了。不错，2头山羊正好是4头山羊的1/2，但是，1/2个蛋糕在现实世界中只能是一个近似值——因为我们需要将某个物品一分为二，而这分成的两个部分可以非常相似，但绝不可能一模一样。

在分割蛋糕等物品时，我们会不由自主地想到分数。在空间测量时，分数同样非常重要。在划分土地或者测量建筑用石块的大小时，我们可以用整数加上测量单位来表示。最初，人们是用身体的某个部位来做测量单位的，例如拇指、脚、肘的长度（肘长指肘部至中指指尖的距离），以及步（passum）。步的1 000倍就是千步（mille passus），即1英里。但是，同数山羊不同，在测量石块大小时，我们有时候用了7次拇指之后，长度还会剩下一点儿，需要使用拇指的一部分来完成整个测量活动。这一部分的长度在0与一整根拇指之间，因此需要使用分数的概念。

尽管希腊人把分数视为与整数不同的数字，但是他们的抽象化实现得并不彻底，因为他们使用分数的方式与我们不同。首先，他们用来表示分数的符号与我们不同。例如，在表示1/4时，他们先直接写出字母δ，用来表示数字4，再在这个本来就容易混淆的字母上方添加一个类似于重音符号的标记。第二个希腊字母是β，因此用β表示2是可以理解的。但是，令人摸不着头脑的是，因为某些不明确的历史原因，

他们在β上方添加一条横线，用来表示2/3。这种奇怪的表示法很可能是从埃及人的习惯做法演变而来的。在埃及人知道的分数中，大多数的分子都是1，而2/3是其中一个例外，因此他们用专门的符号来表示这个分数。

古希腊人用来表示1/2的专用符号也很奇怪——它不是希腊字母，而是一个闪电形状的符号（这个符号也不是标准写法，1/2还有其他几种表示方法）。正因为如此，分数运算实在是一个令人头疼的任务。现代人可以将分子、分母同时乘以一个数，将分数变成方便运算的形式，但是古希腊人没有这样的方法可以利用。他们最常见的做法是买一本加法表，从上面可以查询到1/2 + 1/6的答案是4/6（或者2/3）。尽管古希腊人把分数看成整数的比，但是他们的分数没有明确的分子和分母。

然而，在考虑古希腊人的方法时，我们也不应该认为他们的方法过于简单。对于分数运算可能带来的错综复杂的后果，古希腊人有一定的认识。比毕达哥拉斯晚出生几十年的哲学家芝诺（Zeno）根据分数相加时的奇异特性，提出了一个悖论，他本人也因此享有盛名。芝诺属于埃利亚学派（位于埃利亚，这座古城的旧址就在现在意大利的韦利亚海堡外），该学派认为变化与运动是一种错觉。尽管他们对此深信不疑，但是经验似乎表明变化无处不在。为了捍卫学派的论断，芝诺提出了一系列悖论，以证明我们对变化及运动的认识是错误的。在埃利亚学派看来，这个结果意味着大多数人赖以理解变化的经验是错误的。

在芝诺悖论中，最著名的可能是阿喀琉斯与乌龟赛跑的悖论。阿喀琉斯是芝诺那个时代的超级名人，与他赛跑的却是一只动作缓慢且笨重的乌龟。这显然是不公平的，但是芝诺断言，从阿喀琉斯的角度稍加考虑，就会发现这位英雄追不上慢吞吞的乌龟，条件是阿喀琉斯让乌龟先跑一两分钟。毕竟他是一名英雄，让乌龟占这样的便宜并不

过分。但是，芝诺指出，即使阿喀琉斯跑得比乌龟快得多，他也不可能超过乌龟。

为了方便理解芝诺的理由，我们假设阿喀琉斯的速度是乌龟的两倍。实际上，他的速度肯定要更快，但是这个假设可以简化我们的数学推理，而且无论阿喀琉斯跑得有多快，这个证明过程都同样有效。乌龟起跑之后，我们的英雄还在等待。等乌龟跑出去1码（约为0.9米）之后，他开始追赶。很快，他就跑完了1码的距离，但是，在他跑这段距离的同时，乌龟还在继续往前跑。这段时间里，乌龟将前进1/2码。阿喀琉斯跑完这1/2码后，发现乌龟还领先他1/4码，他不得不继续追赶。于是，这个过程不断重复，永远不会结束。阿喀琉斯每次到达乌龟先前所在的位置时，乌龟已经又前进了一小段距离。因此，阿喀琉斯永远也追不上乌龟。

我们不清楚芝诺是否真的不明白其中的道理，但是古希腊数学完全可以解释这个奇怪的悖论。事实上，由于古希腊人理解分数的方式非常独特（从我们现代人的角度来看），而且他们研究数学时使用了一种非常直观的方法（他们对于几何学的热情经久不衰），所以对于他们来说，这个问题很好解释。古希腊人对整数有一种难以割舍的感情，在他们的心目中，“2”（用字母β表示）的意思是“两个单位组成的集合”。这个概念难以表述，但是与我们现在的理解存在微妙的差别。

他们对分数的理解与我们更加不同。在希腊人眼中，构成分数的两个整数仍然保持着各自的意义，他们把1/2、1/3分别理解成“第二部分”、“第三部分”。我们把1/2理解成一个对象（1）被分割成2个部分，而在古希腊人的眼中，它则是放到一起可以形成1的两个完整对象（2个部分）。就这样，他们彻底回避了1/2带来的抽象化问题。例如，他们不会想到半块蛋糕的近似值问题，因为他们想到的是，两块完整的蛋糕放在一起之后，变成了一块更大的蛋糕。阿喀琉斯与乌龟之间的每一段距离，利用我们现代人使用的算式来表示的话，就是：

$$1 + 1/2 + 1/4 + 1/8 + 1/16 + \ldots$$

而用古希腊人的方法来考虑的话，就是一个箱子被一块单位体积的石头占据了一半空间。然后，我们再装进去一块石头。第二块石头是单位体积的1/2，也就是说，两块这样的石头合在一起正好是一块单位体积的石头。第三次装进去的石头是单位体积的1/4，也就是说，4块这样的石头加到一起才等于一块单位体积的石头。以此类推，这些石头越加越多，箱子越装越满，但是永远也装不满。这是毕达哥拉斯学派的整数情结留给我们的遗产。古希腊人对分数的理解与我们不同，在他们的心目中，分数表示用2个、3个或者4个完整物体拼凑成另外一个物体。

现在，我们知道下面这个级数趋近于2：

$$1 + 1/2 + 1/4 + 1/8 + 1/16 + \ldots$$

每加上一项之后，该级数就会更加接近2，但是永远不能达到2。在理论上，如果这个级数包含该序列的整个无限集合，最终的和就会等于2。但是，无论有多少项，和都不会超过2。我们说，随着级数的项数趋近无穷大，和就会趋近2。然而，在现实世界中，没等乌龟跑出2码的距离，阿喀琉斯就已超过乌龟了。也就是说，芝诺悖论被破解了。毫无疑问，古希腊人的级数求和方法非常直观，比我们的现代方法更容易理解，因为我们可以清楚地看出，由于装填的石头越来越小，箱子永远都装不满。但是，仅从下面这个级数

$$1 + 1/2 + 1/4 + 1/8 + 1/16 + \ldots$$

你却不容易看出它的和是一个有限数。事实上，下面这个极其简单的级数

$$1 + 1/2 + 1/3 + 1/4 + 1/5 + \ldots$$

就不具有数学家所说的“收敛性”。随着项数趋近无穷大，它的和也将趋于无穷大。

接下来，我继续讲关于毕达哥拉斯学派的故事。他们正在考虑如何用整数比的形式来表示2的平方根。如果你对分数略有了解，就会知道这个值与707/500相近，但是又不完全相等。严格地说，在古希腊人看来，这个数应该等于707个1/500。但是，古希腊人可能会把它表示成1、1/5、1/5、1/100、1/500、1/500这种形式，尽管这种表达方式更令人困惑。在绞尽脑汁之后，毕达哥拉斯学派借助逻辑分析，证明$\sqrt{2}$不可能表示成任何两个整数之比的形式，任何整数都不行。这个新数与他们的世界观格格不入，他们认为整数是宇宙基础的信念摇摇欲坠。

证明$\sqrt{2}$不能表示成整数之比，似乎并不是一件容易的事。乍看之下，古希腊人似乎必须逐个检查所有分数，以确保任何分数的值都不等于$\sqrt{2}$。这显然是一件不可能完成的任务，因为分数有无穷多个。但对于酷爱逻辑的古希腊人而言，这不是难事，他们利用自己对奇数和偶数特性的粗浅认识，再加上一点儿逻辑推理，就完成了这项任务。他们采用的是一种常用的经典证明法：首先假设某个命题是真实的，然后根据这个假设得出一个不可能成立的结论，从而证明这个命题是假的。下面是具体的证明过程。

假设某个整数之比（借用古希腊人的习惯，我们设这个整数之比为α/β）可以表示该对角线的长度，也就是说，α、β为整数，且$\alpha/\beta=\sqrt{2}$。为简单起见，我们假设α和β是可以得出这个比值的最小整数，也就是说，这两个整数可以构成2/3这种形式的比，而不是4/6或200/300这种形式的比。

为了避免棘手的除法运算，我们采用了一种老办法：等式两边同时乘以β。由此得到：

$$\alpha=\beta\times\sqrt{2}$$

接下来，左右两边进行平方运算，以消去平方根：

$$\alpha^2=\beta^2\times 2$$

到目前为止，我们使用的都是中学学过的简单数学知识（古希腊人的做法略有不同，但结果是一样的）。古希腊人知道，任何数乘以2，积都是偶数。因此，上述等式的右边是一个偶数，这就意味着α^2也是一个偶数。由此可知，α是偶数，因为奇数的平方肯定是奇数。

所有的偶数都可以被2整除，因此α^2可以被4整除。这就说明$\beta^2\times 2$也可以被4整除，从而证明β^2可以被2整除，也就是说，β^2是偶数。既然β^2是偶数，那么β也一定是偶数，也就是说，它可以被2整除。

我们的目的就要达到了。从上述证明可知，α和β都可以被2整除，因此它们就不可能像最初假设的那样，是比值等于$\sqrt{2}$的最小整数。也就是说，我们经过证明发现，最小的两个整数并不是最小，从而说明原命题在逻辑上不成立。因此，我们可以确定比值等于$\sqrt{2}$的整数是不存在的。

对于现代人而言，$\sqrt{2}$根本不是什么问题，我们可以若无其事地使用$\sqrt{2}$这个数，也可以把这条对角线的长度表示成小数形式，还可以根据需要保留小数的位数，例如，1.414 213 562…。但是，对于有整数情结的毕达哥拉斯学派而言，他们没有任何选择。现在，我们把这样的数称作无理数，原因是它不能表示成整数之比的形式。而对毕达哥拉斯学派而言，这些数似乎真的对理性基础构成了威胁，至少在他们将数字与几何图形联系到一起时，会让他们遭遇麻烦。据说，可怜的希帕索斯因为泄露无理数的秘密而惨遭报复，被扔到水里淹死了。实际上，当时的几何学研究完全不同于数字研究。在毕达哥拉斯

学派的心目中，两者之间的区别非常大，几何图形似乎与数字毫不相干。他们认为，几何学是一种直观概念，而不是一个数字概念。

现在，在提到这些除整数以外的数时，无论是像$\sqrt{2}$这种无理数，还是表示成十进制序列的有理数，例如把2/3写成0.666 666…的形式，我们都可以称之为“实数”。我们可以得心应手地使用无理数乃至所有实数。手动计算已经在很大程度上被计算机取而代之，事实证明，实数的处理毫无难度可言。人们通常根据设备的计算能力，对这些数字进行四舍五入的处理，因此0.666 666…（其中“…”表示该数可以无休止地写下去）就会在最后一位小数上进行四舍五入，变成0.666 667。

总的来说，我们不清楚实数是不是“真实”的——考虑到它与宇宙的本质直接相关，而不是一个数学概念。在数学中，圆的周长与直径的比值是无理数π，但是在现实世界中，我们从来没有也不可能完美地实现这个比值。即使我们通过某种方法画出一个完美的圆，但是鉴于它是由原子构成的，我们仍然需要考虑粒度问题。无论物质世界之中的这个圆是如何构造的，无论它是一块圆形金属还是纸上画出来的图形，都不会具有完美的连续性。这就意味着我们永远无法在我们构建的实物与数学的预言之间建立理想的一一对应关系。

当然，我们可以将物质排除在外，只考虑空间自身的问题。但是，现代物理学告诉我们，即使空间也可能存在粒度问题，并不是完全连续的，至多只能在普朗克长度这个量级实现量子化。普朗克长度是不可测量的极小长度，约等于10^{-35}米，是最小的原子——氢原子直径的10的25次方分之一。然而，现实世界的这种不可避免的粗糙性在数学世界中却根本不存在。在那里，一切都有可能，数值的连续性更是惯常现象。宇宙可能无边无际，某些量子特性可能与实数保持一致，但这仅仅是一种可能，而实数很有可能只是现实世界的一种近似表示。

然而，古希腊数学家并不了解物质世界的这种特性。他们没有花费太多精力去寻找一种可以与现实世界精确吻合的形式数学，而是一头扎进理论数学的研究之中。在这个与尘世隔绝的完美世界里，有一个神秘的代表人物——欧几里得。

---

1. 亲和数指一对存在特殊关系的数。一般来讲，如果两个数中任何一个数都是另一个数的真因数之和，则称这两个数是亲和数。——译者注
2. 有的数学家认为，所有数字都非常重要。但如果你认为某些数值小的数字不那么重要，那么这些数字也会因为小而变得重要。
3. 1英里≈1.61千米。——编者注

第4章
欧几里得：几何定理的完美证明

Are Numbers Real?

上学时，如果你接触过几何学，并且在解题结束时以胜利者的姿态（或者怒气冲冲地）写下QED（拉丁语“quod erat demonstrandum”的缩写，意为“证明完毕”）这个表示几何证明过程结束的传统标志，就说明你正走在欧几里得所指引的道路上。在大约2 000年的时间里，他的著作《几何原本》一直是至高无上的数学教科书。但是，关于欧几里得本人，我们却知之甚少。有的学者甚至认为根本不存在这样一个人，欧几里得只是代表若干作者的一个虚构的姓名。我们仅知道他的著作出自托勒密一世在亚历山大创建的那所无与伦比的学校，欧几里得（如果他确实存在）是这所学校的一个重要人物。欧几里得是当时公认的名师，他的那部代表作是他授课所用的教科书。在撰写这部书时，他没有打算收录那些惊天动地的新发现，而是收集了那些已有的知识。然而，这些知识的编排方式却给人留下了极其深刻的印象。

欧几里得几何学建造了一个自得其乐的小小世界，从此以后，数学研究就沿用了这个模式。《几何原本》首先提出若干假设，然后在此基础上完成一系列逻辑清晰的证明。在证明的过程中，人们不需要从事诸如观察、计数、测量等苦活儿去解决现实世界的难题。现在，我们把这部著作看作一本权威的几何教科书，但实际上，书中还涉及

算术，也介绍了古希腊人对代数问题的直观表示。从级数1 + 1/2 + 1/4 +1/8 + ...这样的直观表示可以看出，古希腊人已经开始接触代数学了。

毫无疑问，证明的概念与精确性（与测量人员工作时只求大概的结果不同）是古希腊人对数学的最大贡献。我们知道，古埃及人在几何学的舞台上非常活跃，古巴比伦人的数字与代数知识也达到了古希腊人无法企及的高度。但是，这两个时间更早的文明对证明之道却不怎么关心，因为他们认为有数字或者图形就足够了。他们研究的仅仅是理论数学。

我们已经接触过数学证明的概念。数学证明始于毕达哥拉斯学派，但是欧几里得为证明过程建立了一套严格的模式。他首先为证明规定了一系列条件、公理或者“已知内容”。这些条件、公理或者“已知内容”无法加以证实，但是数学家要完成证明，就必须视其为理所当然。这些公理都是显而易见的（尽管有个别例外，但基本上确实如此），可以用作证明过程的基础。证明时，数学家必须从这些公理出发，环环相扣地完成一系列逻辑推理步骤，直到最终完成证明。

欧几里得的《几何原本》（从技术上看，应该称之为一套书，因为每卷的实际长度都受到限制）首先给出了一些定义（包括点、直线、圆的定义），然后是5条公设（也被称作“假设”）和5条公理（现在也被视为“公设”）。紧接着，他给出了第一条定理，描述了利用圆规和直尺作等边三角形的方法（古希腊人尤其热衷于尺规作图）。事实上，欧几里得给出的一些定义并不严谨缜密（例如，他用“倾斜角”这个名词来定义角，但实际上，“倾斜角”这个词反而比“角”更加冷僻），但是他所表述的意义都是正确的。本书给出这些公理，但不再提供详细的证明过程。

5条公设是：

1. 过任意两点都可作一条直线；

2. 一条直线可以无限延长；

3. 以任一点为圆心，任意长为半径，可以作一个圆；

4. 所有直角都相等；

5. 一条直线和另外两条直线相交，若两个内角之和小于180°，则这两条直线一定相交。

5条公理是：

1. 等于同量的量彼此相等；

2. 等量加等量，其和仍相等；

3. 等量减等量，其差仍相等；

4. 彼此能重合的物体是全等的；

5. 整体大于部分。

这些公设和公理大多是显而易见的常识，但是数学证明的严谨性要求我们不得进行不必要的假设。证明过程不接受常识作为论据。在数学世界中，我们可以做出在现实世界中不成立的假设，这是数学世界特有的有利条件之一。例如，我们从定义中可以看出，直线“只有长度，没有宽度”。而在现实世界中，线条肯定是有宽度的。直线的定义把我们置身于一个无法真实存在的世界之中。

此外，至少有一条公设对数学特有的环境做出了若干假设。这些假设没有出现在公理之中，是因为欧几里得从未想过这方面的问题。欧几里得的第五公设（该公设的意思是不平行的直线就会相交，但其语言表达异常复杂，它暗藏的意思就是平行的直线不会相交）在欧几里得研究的平面上的确是正确的，但对于在现实世界中更加常见的曲

面，这条公设却不一定正确。例如，我们沿着朝北的方向，作若干条与地球赤道垂直的直线。这些直线就是地球的经线，在它们到达北极时就会相交，但是在赤道这个位置上，它们确实是互相平行的直线。所以，第五公设在平面上是成立的，但是在曲面上却不适用。

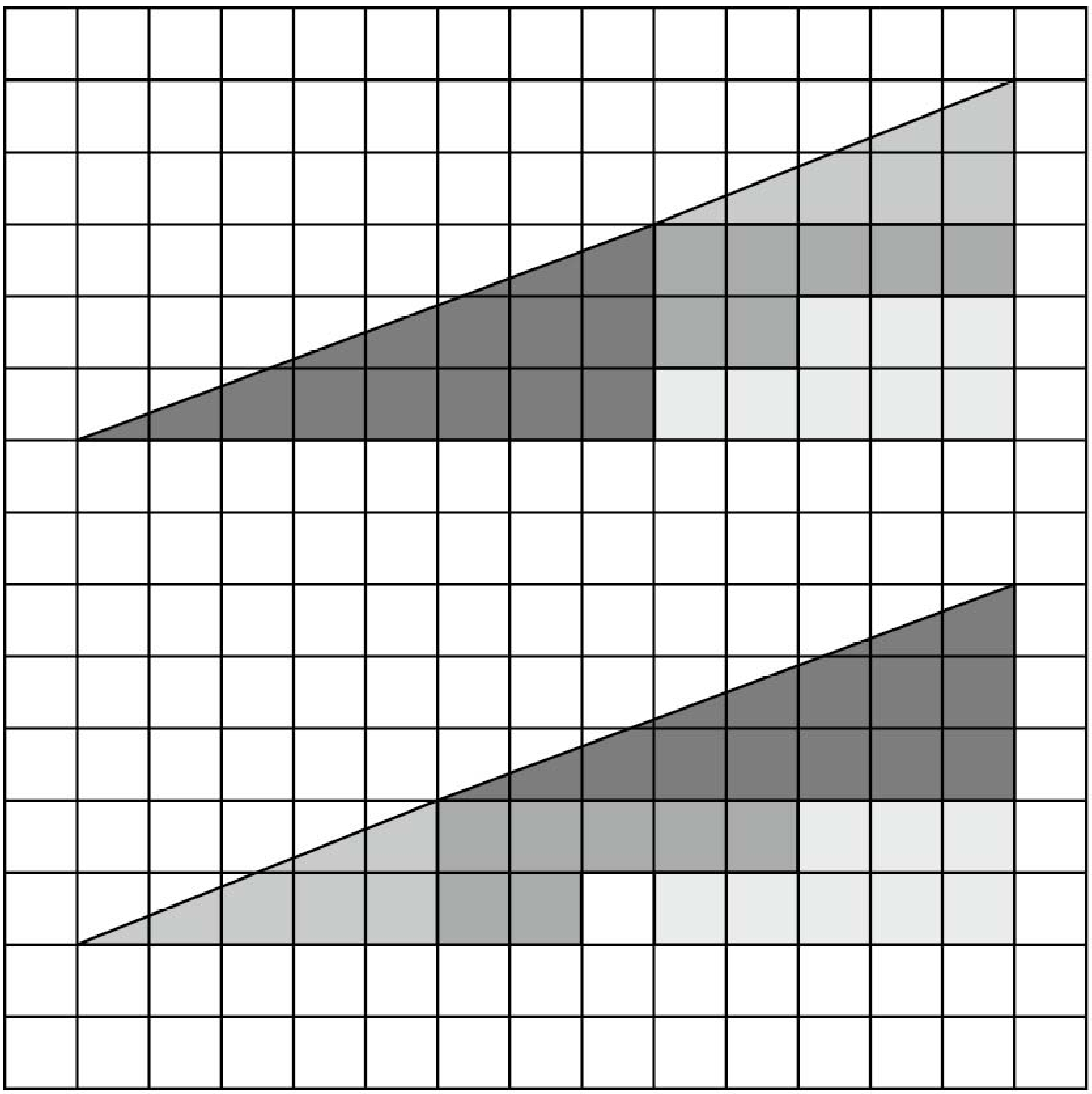
欧几里得几何学必须具备的一个基本条件是平行线不相交，否则它就会轰然倒塌。直到19世纪，人们才提出了适用于现实世界中的曲面的非欧几何。当时，除了测量地球表面以外，非欧几何的实用性似乎十分有限。但是，后来的事实证明，在爱因斯坦开始研究广义相对论时，非欧几何的研究领域已经延伸至多维曲面，并成为一种非常重要的研究工具。

想到数学可以彻底摆脱我们所在的物质世界继续存在并发挥作用，我们不禁要问，欧几里得的研究对象是现实世界吗？答案是否定的。但是，这并不说明他的研究就毫无价值。欧几里得几何学为现实世界绘制了一幅逼真的画像，非常珍贵，但它描绘的却不是一个真实的世界。也许最能说明欧几里得定理与现实世界具体对象之间关系的就是柏拉图的那个隐喻。这位成名时间比欧几里得早几百年的哲学家设想了一个犹如天堂般完美的宇宙，在这个世界里，欧几里得研究的那些数学概念是可以存在的。这些完美的图形和物体具有某种超现实性，而我们在现实世界里感知的那些图形和结构仅仅是这些纯净原型的投影。

具体来说，柏拉图用投射在山洞中的影子打比方。在他眼中，不完美的三角形就是在纯正数学世界的强光照射下，一个完美三角形留在山洞中的模糊阴影。他的这种理解有一定道理。这并不是说数学从某种意义上讲比现实世界更加真实或者更加完美，而是说宇宙过于复杂，我们很难将它惟妙惟肖地表现出来。因此，数学通常只能是真实宇宙的一种近似表现。整数在现实世界可以找到完全对应的对象，而分数则不可以。同样，现实世界是由原子和线条构成的，这些线条不

仅有宽度，而且不是毫无瑕疵的直线，因此几何图形也不是现实世界的完美表现。

洞中影子的不完美特点（数学证明中的完美图形与我们在现实生活中处理、体验的那些对象之间的区别），在“失踪的正方形”这个令人费解的难题上得到了充分体现。



在这幅图中，两个图形看起来一模一样，只是它们各部分所处的位置不同。可以看出，下面的图形中有一个空白格，这说明下面图形

与上面图形的面积并不相同。柏拉图的完美图形是不会出现这种情况的，但是在模糊、朦胧的现实世界中，画出来的线条都不精确，因此我们难免遇到这种情况。在上面的图形中，较小三角形的左边的角与较大三角形的左边的角看起来度数相同，但实际上并不相同。重新排列之后，那个空白格就是这个极其微小的差别造成的。

随着不断深入研究数学与现实世界的相互关系，我们越发清楚地看到在科学研究中使用模型的意义。模型就像在现实基础上制作的玩具。宇宙是一个庞大的存在，即使其中一个很小的部分（例如人）也有非常复杂的结构。在了解宇宙奥秘、确定其物理构成时，我们通常需要做简化处理，降低研究对象的复杂程度。

模型可以是看得见摸得着的具体表现，例如太阳仪。在计算机问世之前，这种简洁的太阳系机械模型一直深受欢迎。然而，我们使用的模型大多是由一系列方程或者数学系统构成的“数学模型”。它们具有现实世界中物体的某些特征，但是处理起来要简单得多。现在，人们通常会通过计算机编程来建立这样的模型，尽管很多科研人员仍然更青睐以简单（或比较简单的）方程式为核心的模型。

我们将在后续章节深入讨论这些模型。从本质上看，它们表示的都是柏拉图理想世界中的宇宙的整体或部分情况。它们不是真实的事物，而是真实事物简单化、完善化的产物。有时，这些模型被称作“原型”，这个名称使我们不由得想起中世纪对柏拉图思想的美妙描述——“类型和影子”。只不过，柏拉图把相互关系弄颠倒了。在他的想象中，先有真实存在的完美对象，例如三角形，然后才有我们在日常生活中遇到的有瑕疵的影子。根据柏拉图的理解，我们周围世界的真实程度比不上理想世界。但是，科学家建立模型时，情况则正好相反，有瑕疵、有限制条件的是他们建立的模型。科学模型与科学理论是自然世界的数学投影，比真实系统要简单得多。

柏拉图的山洞本身就是一个模型，它毫无疑问是山洞外世界的一个表现，但这不能说明任何问题。你也可以说数学就是大家共同创作的一部科幻小说，是一个所有数学家都愿意和谐共处的精神世界。但是，这个世界不允许存在科幻小说的天马行空，这里的所有规则都必须明文规定并得到一致认可。任何数学内容，只要遵从了这些规则，就会得到认可，无论它是否与现实世界存在相似之处。但是，经常有人提到一个事实：很多数学内容不仅与现实世界有相似之处，而且在反映现实世界真实情况这个方面具有异乎寻常的能力。其中的原因可能仅仅在于数学的基本结构中有很大一部分（例如整数运算的本质）是基于对现实的反映，但是，在我们摸清了数学世界的真实情况后，我们就会重新思考是否有其他原因。

即使是在完美图形这个壁垒森严的世界里，欧几里得的门徒们也在化圆为方的道路上遭遇了滑铁卢。他们的内心有一个情结，想凭借一个圆规和一把直尺画出世间万物。他们想，是否可以画出一个正方形，它的面积恰好等于圆这个最简单、最完美的图形？这个想法令古希腊人难以释怀，他们甚至赋予那些埋头钻研这个难题的人一个专门的称呼——“化圆为方者”（tetragonidzein）。尽管这个称呼听起来很威风，但他们的努力不过是一种不自量力的行为。

现代人都知道，在计算圆的准确面积时，古希腊人遭遇了另一个比$\sqrt{2}$更重要的无理数——圆周率（π）。圆周率不仅是无理数，而且是超越数，也就是说，我们无法通过一个有限方程得出它的精确值。（我们可以用方程表示圆周率的值，但是这些方程都是无穷级数，因此，我们甚至无法用方程来表示圆周率的精确值。）

在欧几里得那个时代，数学家在面对几何学的严谨性时感到心安理得，我们也可以肯定，他们从事数学研究的目的并不单纯是为了刺激智力发展，与此同时，数学还得到了广泛的应用。我们已经知道，“几何”（geometry）这个词源于希腊语，原意是指测量地球。无

论你是测量人员，还是建筑师，对于你来说，测量地球的近似值都具有非常重要的意义。但是，推导和证明却是在有条不紊、与世隔绝的数学世界里完成的。与这个世界相比，我们身处的现实世界混乱不堪，令人气馁。

杂乱无章的物质世界似乎是欧几里得精确证明的墓地。但是，当一位数学家登上历史的舞台，甘愿干苦活累活，并将数学应用到工程技术乃至战争武器等领域时，完美世界的壁垒被打破了！

第5章
阿基米德：用沙粒填满宇宙

这位为当时的数学研究注入实用性的古希腊数学家就是阿基米德。不仅如此，他在非实用性数学方面的研究同样了不起，水平之高令当时的大多数人都感到不可思议。

尽管后世的翻译家成功地复原了不少古希腊文本，但还是有很多遗失了，其中就包括很多重要人物的生平传记。我们知道，阿基米德去世之后，曾经有人为他立传，但是这部传记却没有流传下来，所以我们无从知晓他的具体生平，只知道他大约出生于公元前287年。阿基米德不仅具有不逊于先贤的几何天赋，毫无疑问，他还是人类历史上的第一位工程师，因为他把数学应用于实践，发明了武器和实用的机器。

毕达哥拉斯学派曾经凭借想象，在底层建构上将数学与天地万物联系在一起。在阿基米德时代，人类不可能回归纯粹的数学，但是阿基米德成功地证明了数学是科学发展的必需品，它可以成为自然哲学家的忠实仆人。有人甚至认为，阿基米德是第一位现代意义上的科学家。然而，在阿基米德本人看来，他只是一位应用数学家。希腊历史

学家、传记作家普鲁塔克说，阿基米德认为自己的机械发明不过是“研究几何学之余的消遣”，根本不值一提。

最值得我们关注的是，阿基米德帮助数学在另一个方面摆脱了与现实的联系。我们知道，古希腊数学（具体来说就是古希腊几何学）有很强的直观性，它永远不可能为数学插上想象的翅膀，也不可能构建起我们现在看到的梦幻一般的数学王国，原因之一就在于古希腊的数字系统具有很大的局限性。古希腊数学几乎从未采用现代人熟悉的符号系统，也没有想办法降低数字系统的复杂程度。我们知道，古希腊数字都是字母表中的字母，为了扩大表示范围，还选用了几个古代字母。因此，这些数字使用起来非常不方便。最令人感到气馁的是，这套数字系统到10 000就结束了——米亚德（10 000）是最大的数字。

为了证明数字系统可以发挥更强大的功能，阿基米德撰写了一本名叫《数沙者》（*The Sand-reckoner*）的小册子。在这部不同寻常的著作中，他计算了填满整个宇宙需要的沙粒数量。显然，他在实践性工程技术方面取得的成就并不包括这项计算。他的目的似乎是证明数学不会受到世俗思想的束缚，也不会因为当时采用的数字系统而裹足不前。为了达到这个目的，阿基米德必须另辟蹊径，发明一套全新的数字系统。

为了增加权威性，阿基米德把这本书献给了叙拉古的统治者——革隆。阿基米德似乎受雇于革隆，甚至有人认为，生活在叙拉古这个动荡不安的小地方的阿基米德和革隆十分投契。阿基米德告诫革隆不要错误地认为地球上的沙粒无法数清，或者以为沙粒的数量为无穷多。他还明确地告诉革隆，他所说的沙粒“不仅指叙拉古附近和西西里岛其余地方的沙粒，还包括地球上所有其他地方的沙粒，无论有没有人居住”。但是，之后他把地球上的沙粒数量这个问题放到一边，着手证明他可以找到一个数，用来表示填满宇宙所需的沙粒数量。

以现代人的眼光来看，这项计算任务是不可能得到一个有意义的结果的，因为我们根本不知道宇宙到底有多大。我们仅知道可观测宇宙的大小。这是由光在宇宙诞生以来可以传播的距离决定的，大约为900亿光年。在这个距离之外，我们就一无所知了。然而，在阿基米德那个时代，宇宙的概念则简单得多。在亚里士多德的简单模型中，宇宙的中心位置不是太阳，而是地球——地球是世间万物的中心。就像我们每日所见的那样，天空中可以观测到的所有东西，包括月亮、太阳和各大行星，都在围绕地球运转。天空的外面是一个球体，上面点缀着一些稀疏的光点——恒星。按照这个推算结果，整个宇宙比我们现在已知的太阳系还要小（要知道，根据古希腊人的宇宙模型，恒星位于土星轨道的外侧，因为土星是当时已知最遥远的行星）。

有意思的是，阿基米德在《数沙者》中明确提出了另外一种观点。据他介绍，萨摩斯的阿利斯塔克在其著作中指出，太阳与恒星都是静止不动的，地球绕着太阳转，太阳位于宇宙的中心。这些只言片语引起了人们的兴趣。阿利斯塔克的那部著作没有流传下来，因此我们不知道阿利斯塔克到底说了什么，只能根据若干间接引用，推测这个地球围绕太阳运转理论的首创者的相关信息。

阿利斯塔克说，他心目中的宇宙远远大于当时人们普遍采用的模型。阿基米德指出，阿利斯塔克描述的宇宙大小令人困惑，根本不可能是正确的。他的理由是，阿利斯塔克似乎在书中说过，镶嵌有固定恒星的天球十分遥远，地球的运转轨道与天球相比，就相当于“天球的中心与它的表面相比”。阿基米德指出，天球的中心没有大小，无法与其表面构成比率关系。但是，阿基米德认为他知道阿利斯塔克这番话的真实含义——这是在拿地球的大小（地球是传统宇宙模型的中心）与天球的大小做比较。

在明确这一点之后，阿基米德开始计算填满传统模型代表的宇宙和阿利斯塔克模型代表的体积更大的宇宙所需要的沙粒数量。首先，

他必须计算出这两个宇宙到底有多大。作为一名古希腊人，阿基米德在计算中自然会大量使用几何知识。开始计算之前，他对地球的大小以及地球、月亮和太阳的相对大小做出了一系列假设。大多数的假设都比较直截了当，例如，“地球的周长最多约为3 000 000斯塔德”。［斯塔德（stade）是根据体育场跑道长度确定的单位。］但是，有一个假设比较难以理解：“太阳的直径大于宇宙天球的最大切面的内接千边形的边长。”

有必要告诉大家，千边形（这个词在现代并不常见）是指有1 000条边的多边形。阿基米德在这里对“宇宙天球”这个概念下了明确的定义。我们往往以为希腊模型中宇宙的尽头是恒星，但是阿基米德指出，“宇宙天球的中心就是地球的中心，它的半径等于太阳中心与地球中心的直线距离”。换句话说，他认为“宇宙天球”就是从地球上看到的太阳运行轨道所构成的球体。也就是说，他认为太阳的直径大约是我们观察到的太阳轨道的千分之一。

在进行科研活动时，古希腊最流行的做法是纯粹依靠论证，而往往不屑于验证论证结果与现实之间是否存在相似之处。然而，天文学是个特例。阿基米德绝不是一位只会空谈的哲学家，而且他的父亲生前是一位非常活跃的天文学家。太阳大小与运行轨道的比值，就是阿基米德通过有限的天文学实践活动得出的。他准备了一根长竿，上面安装了一个可以调节位置的圆盘。然后，在太阳刚刚升起的时候（因为他相信，在太阳刚刚升起的时候观察太阳是安全的。其实，他错了，虽然这时候的阳光确实弱一些，因为阳光在到达地球之前，要在空气中传播一段很长的距离），他将长竿指向太阳，然后上下移动圆盘，使它正好遮住太阳。他计算出，当圆盘盖住太阳时，圆盘与视线之间的夹角介于直角的1/200和1/164之间。根据这个结果，他计算出太阳直径与太阳运行轨道的比值。最后，通过取整，他得出太阳直径约为太阳运转轨道的千分之一这个近似值。

阿基米德计算距离时使用的长度单位是斯塔德（源于希腊语的“stadion”一词，复数形式为“stadia”），这给我们造成了不小的麻烦。前面说过，这个长度单位是根据运动场跑道的长度来确定的，但它并不是一个标准单位。1斯塔德可能等于600希腊尺，但是具体数值因地而异，变化范围大概为150—200米（490—650英尺）。然而，即使这个定义非常模糊，阿基米德的计算结果也表明宇宙的外围达到了带外行星的位置。考虑到古希腊人的估算水平，这个结果已经相当不错了。

接着，他武断地认为只需要1米亚德（10 000）粒沙子就可以填满一粒罂粟籽，而罂粟籽至少是手指宽度（比尺小的标准长度单位）的1/40。现在，他需要想办法建立超过米亚德的数量级。因此，他创造了一个新数——米亚德米亚德（亿），并把达到这个值的数称作第二级单位。然后，他通过求米亚德米亚德的平方得到第三级单位，再通过求米亚德米亚德的三次方得到第四级单位，以此类推。在求得米亚德米亚德的米亚德米亚德次方之后，第一期（period）结束，然后他以同样方式得到了第二期。

在建立了这些数量级之后，阿基米德又制定了一系列运算法则。最后，他计算出传统模型代表的宇宙可以装下的沙粒数不超过1 000个第七级单位（即10^{51}，也就是1后面有51个0）。而亚里士多德模型所代表的那个更大的宇宙可以装下的沙粒数为1 000万个第八级单位（即10^{63}）。其实，阿基米德发明的数字系统并没有这么先进，它是以古希腊人当时使用的那套糟糕的数字系统为基础的，所以使用起来十分不便。然而，在计算宇宙大小的时候，尽管阿基米德使用的是几何学知识，但在方法上却有一种别出心裁的新意。

毫无疑问，这个新方法帮助阿基米德完成了计算，也帮助他从同时代的古希腊人当中脱颖而出，从一个全新的角度推动了数学的抽象化。他研究的课题可能是数沙子，但他发明的数字系统摆脱了传统方

法的羁绊，在将数字与实物分离开的道路上迈进了一步。尽管阿基米德迈出了这一步，但他的数字系统仍然缺少一个至关重要的组成部分。当时的人都没有发现，他们的数字系统真正缺少的正是“什么都没有”这个概念。

这是一个深不见底的漏洞！

第6章
斐波那奇：阿拉伯数字的登场

Are Numbers Real?

让古希腊人和古罗马人烦恼不已的一个最重要的原因是，他们的算术中缺少了一个非常重要的元素——0，以至于他们没有办法表示“无”这个概念。从某种意义上看，即使是那些在兽骨或陶片上记账的早期“会计师”，也已经有了“无”这个概念。他们用没有任何刻痕、光溜溜的陶片来表示某种东西不存在。也许我没有办法把“没有山羊”这个概念直接表现出来，但是我可以展示看不到一头山羊的空荡荡的牧场；我也可以拿出一只空盒子，告诉你里面没有橙子。当我们有了“无”这个概念之后，在不知不觉间，我们就不再只用一个数字专门表示数量与实物之间的对应关系，而是用到了容器。

然而，从表示没有某种实物的“无”，到在数学中十分活跃的0，这是概念上的一大进步。如果没有0，数学家就不可能将数字真正纳入自己的掌控之中。此外，这个神奇的数字还可以方便地充当占位符的角色，增强书面数字的条理性，使人类有可能完成之前根本不可能完成的复杂的计算。

“0”看上去无足轻重，貌不惊人，但是它的出现推动了数学的发展。在相当长的时间里，人们都认为0是一个特殊的数字，很多数学家

甚至认为0根本就不是数字。其中的原因不难理解。在进行一些基本的算术运算时，0往往会导致意外的错误。数学家都不喜欢特例（在这个方面，科学家与数学家的态度是一致的），而0却是最爱出风头的特例。任何数加上0或者减去0，都不会发生任何变化——0是拥有这个特点的唯一整数。任何数乘以0，结果都是0。0就是粉碎机，可以粉碎所有数字。

如果用0做除数，就会得到算术上无法处理的可怕结果——无穷大。我们可以这样想，10除以1，结果是10。那么，10除以1/2呢？在学习分数时，老师告诉我们，除以1/2相当于乘以2，所以10除以1/2的结果是20。（我们也可以这样想：把10块蛋糕分别一分为二，就会得到20块小蛋糕。）同理，10除以1/4，结果是40。随着除数越来越小，结果就会越来越大。当分数的分母趋近0时，分数的值就会趋于无穷大。

然而，这还不是0带来的最大的麻烦。有没有想过0除以0会有什么样的结果？苹果手机的语音控制功能Siri给出的回答令人印象深刻：

> 无法确定。设想你有0块饼干，平均分给0个朋友，每人可以分到多少块饼干呢？看到没有，这个问题没有任何意义。甜饼怪兽很伤心，因为你没有饼干；你也会感到伤心，因为你没有朋友。

分子为0的所有分数都等于0，而分母为0的分数则趋于无穷大。分子与分母同时为0的分数既可以同时归属这两大阵营，也可能被双方同时拒之门外。当印度人第一次使用0这个数字时，数学界就0除以0这个问题的答案展开了大讨论。在一段时期里，这两个看上去似乎都有道理的结果得到了很多人的认可。公元7世纪，婆罗摩笈多[注]认为0除以0应该等于0；公元12世纪，婆什迦罗[注]宣布，0除以0的结果是无穷大。

后来，数学界（也像Siri一样）都赞成0除以0的结果无法确定的说法，认为0除以0没有确定的答案。在数学上，0除以0这个问题并不是一定要得出一个有意义的结果。随着历史的变迁，我们越来越清楚地看到，大多数的数学知识并不是以现实为基础的。在这种情况下，数学家可以随意制定规则。0除以0就符合这个条件。

正是由于0的这些奇异特性，在刚开始的时候，数学界常常认为0不是整数。（后来，数学界决定不把数字1视为素数，也是出于同样的原因。）但是，当数轴这个非常重要的概念出现之后，把0排除在整数之外就会导致一个问题。在过去30年里受过教育的人都会在学校里学到数轴这个概念。如果有人没有学过，我在这里介绍一个非常方便的理解方法。大家可以想象一把两端无限延伸的直尺，上面的每个主要刻度都代表一个整数。沿着直尺向前（朝右），数值越来越大；沿着直尺向后（朝左），数值不断减小。

数轴上既有正数，也有负数。问题是，当数值小于1且朝着负值的方向移动时，会出现什么情况？公元纪年法是人类实践活动中最早出现的数轴之一，它在这个问题上就犯了错误。525年，僧侣狄奥尼修斯·伊希格斯发明了这套公元纪年系统。8世纪30年代，历史学家比德对其进行了推广。到9世纪，这种纪年方法已经得到了基督教国家的普遍认可。在这种纪年系统中，公元元年（AD 1）之前是公元前1年（1 BC），竟然没有公元0年。这种情况直到今天也没有更正过来。

时间线在–1（1 BC）之后就迅速跳到了+1（AD 1），中间没有任何数。AD表示“Anno Domini”，意思是“主的年份”，AD 1指的是耶稣诞生之年（尽管这个说法存在争议）。因为数轴上存在这个缺口，一些立志成为历史学家的人在计算生卒年跨公元前和公元后年份的人物的年龄时往往会出错。

历史学家不同于数学家，这种从–1直接跳到1的纪年方法得到了他们的认可。但是，如果同样的情况出现在数轴上，人们就无法接受

了，因为数轴是一个非常重要的数学工具，在数轴上前进或后退的行为应当与加法或减法运算相一致。例如，我们可以利用数轴来计算5 + 2的得数：从5开始，朝前（正方向）移动两个单位，就会得到答案7。但是，如果数轴上没有数字0，在计算1–1时就无法得出正确答案0，因为从1开始朝后（负方向）移动一个单位，就会到达–1的位置。无论我们是否愿意，若想让算术基本运算可行，整数中就必须包含0。

如果0的作用仅仅是为数轴填补空缺，表示“无”的概念（位于整数–1和1之间，表示两者之间的那个整数），那么虽然这足以让它成为一个非常重要的数字，但还不足以促使数学领域发生翻天覆地的变化。0的重要作用还体现在它的另一个身份——数字中的占位符上。我们已经知道（参见第2章），罗马数字和更早的希腊数字在数位排列上缺少辅助系统，随着数字增大，它们会变成庞大到难以处理的字符串，从而大大增强了计算的难度。但是，在实践活动中，这两大古代文明对一种存在了1 000多年的有效方法却一直视而不见。

这个方法可以追溯至古巴比伦人。古巴比伦人从他们的苏美尔祖先那里获得了灵感，开始使用六十进制的计数系统。这套系统用竖直的刻痕表示1（这与符木刻痕非常相似），用斜向一侧的符号（称作“钩”）表示10。我们把一天分成24个小时，又把每个小时分成60分钟。与之相似，古巴比伦人从1数到10之后，就开始采用六十进制。他们的祖先用不同的符号表示1和10，但是表示60的符号与表示1的符号相同，只是更大一些。用不同的笔可以刻画出更大的符号，但是很容易混淆，因此，后来的古巴比伦人换了一种方法，借助数字的位置来达到这个目的。

用罗马数字写一个大数，例如MDCCCLXXVII，就会发现这个表达法浪费了传递大量信息的机会。虽然罗马数字中字母的位置也会发挥作用，例如LX表示60，而XL则表示40，但总的来说，在这套数字系统中，字母位置的作用十分有限，只表示字母的先后次序。但是，

字母位置传递信息的潜力是巨大的。例如，“GOD”（上帝）和“DOG”（狗）是由相同的字母构成的，但是由于这些字母的排列位置有所不同，因此两个单词的意思也完全不同。数位携带了更多的信息，难道不能更广泛地对其加以应用吗？古巴比伦人正是这样做的。他们将一个数字最右端的数位定义为1，向左的第二个数位表示60，第三个数位表示60×60，以此类推。

也就是说，如果用Y表示古巴比伦人的“1”，用D表示古巴比伦人的“10”，那么YY DY这个数字等于2×60 + 10 + 1，也就是131。由于数字位置表达了更多的信息，因此在表示大数时要简洁得多。不仅如此，采用这个办法之后，我们还可以将类似的数字上下对齐，排成列，从而大大降低加法、减法和乘法等运算的难度。例如，在求YY DY与Y Y的和时，我们可以按照数位对齐，写出下面这道算术题：

YY DY

Y Y

我们立刻就可以看出，答案是YYY DYY。我们甚至没有意识到，这是一道可怕的六十进制算术题。当然，就像我们现在的数字系统一样，列与列之间的进位还需要遵循进位法则。尽管如此，这也比古希腊人和古罗马人使用的数字系统高效得多。古巴比伦人甚至可以处理无理数问题，例如，令毕达哥拉斯学派挠头的2的平方根。同样令人震惊的是，他们的数字系统中也有小数（因为他们使用的是六十进制，所以这些小数其实是六十进制小数），古巴比伦人留下的陶片还给出了2的平方根的近似值：1.414 222。很难理解，古希腊人竟然没有学习古巴比伦人的这些方法，反而后退了一大步，个中原因已经湮没在历史的长河中了。

然而，古巴比伦人的那套数字系统也有一个问题。上面那道题中的第二个数是Y Y，即61。如果这个数是3 601，该怎么表示呢？60×60

的列是Y，60的列中什么也没有，而表示1的那个列中有Y，因此3 601被表示成Y Y。严格地讲，两个Y之间的空格应该稍大一些，但是很难清楚地区分这两个数，在手写陶片时就更难以分辨了。从现存的古巴比伦人陶片来看，这些数字的排列方式极具随意性。当然，在第2章讨论的邻居借山羊的问题中，数清那些山羊并不是难事，因为我不大可能借给邻居3 601头山羊。所以，根据当时的情境，我们可以看出那些数字到底表示什么意思。但是，如果在交易物品时使用的是比较小的单位，或者涉及钱财时，这些数字就难以区分了。

在古巴比伦文明走向衰落、古希腊文明即将兴起的时候，人们找到了一个解决办法。如果某一列中没有数字，他们就在那里画一道斜线，表示这是一个空列。因此，61仍然是Y Y，而3 601则被表示成Y \\ Y的形式。太棒了！按列计算的方式要安全得多，而且人们不需要根据情境来分析数字表示的含义。但是，\\与现代数学中的0还是有所不同的。因为某种原因，古巴比伦人一直没有承认一个事实：由于数字后面有可能出现除号，所以，数字中某一列为0时，仍然有可能与其他数字混淆。无论何种原因，\\都无法充分发挥占位符的作用。这个符号从来没有独自出现过，也没有出现在计算过程中，因此，它没有发挥整数0的作用，而仅仅是0的幻影，只是有时起到占位符的作用。只有身兼二职，这个符号才有可能真正地发挥作用。

古希腊文明末期的某些数字系统也出现过这种偶尔使用占位符的情况。在古希腊文明走向衰落的时候，天文学的研究得到了蓬勃发展，从相关研究成果中就能找到这些占位符。古希腊数字系统使用了十分笨拙的表示法，用单个字母分别表示从1到10的整数、各个小于100的10的倍数以及各个小于1 000的100的倍数，但也有一些系统使用的方法比较接近于古巴比伦人的方法。后来的希腊人同我们一样，用度、分、秒来表示角度。例如，在表示5度0分20秒的角时，希腊人会在分甚至秒的位置上画一个圆圈，并在圆圈上方画一个复杂的棒形标志，以示空缺。

至此，从某种意义上看，他们已经朝着0迈出了一小步，但是这种占位符号仍然没有被他们视为独立的数字。在表示普通数字时，希腊人会在对应的希腊字母上方添加一条横线，但是在空白的占位符上方，他们会添加另外一种标志，以表示这是一个特别的符号。在我们看来，把这种占位符变成一个真正的数字，似乎是一件理所当然的事。尽管在记账时以及在天文学研究的文本中都不可避免要使用数字，但我们也别忘记了，古希腊数学在绝大多数情况下都是几何学研究，非常直观，关注的都是各种图形。正是由于思维上的这种特点，这种表示空缺的占位符在几何学实践活动中是难以想象的，也很难加以利用。

也就是说，即使某些数学家和哲学家非常聪明，有可能提出0这个数学概念，也无法找到合适的出发点。至于那些必须与数字打交道的古希腊会计人员和商人，他们通常会使用一种不穿绳的算盘。这种算盘其实是一块平板，上面装有一排排算盘珠。与我掰手指数山羊的方式相比，算盘稍微先进一点儿。使用算盘时不需要占位符，因为算盘珠在算盘上的位置是固定的，表示十位数、个位数以及其他数的算盘珠都在对应的位置上。如果算盘上的某个位置没有算盘珠，也没有任何问题。但是，这也根本不算一个数字，而是表示这种高级符木上出现了一个空缺。

直到13世纪初，0才以我们现代人常见的形式出现在几位西方数学家的笔下。其中最有名的当属比萨的列昂纳多，但他的另一个名字——斐波那奇却更广为人知。斐波那奇生于1170年，他的身为外交家的父亲代表比萨出使北非时，可能把年轻的斐波那奇也带去了。正是在北非的游历，让他接触到了阿拉伯数学家从印度学习并改进的那套新颖的数字系统。斐波那奇有一本名叫《计算之书》（*Liber Abaci*）的专著。这个书名很奇怪，它的意思是“算盘书”，但是这本书与算盘没有任何关系。通过这本书，斐波那奇不仅把阿拉伯数字（后演变为现在通用的阿拉伯数字）引进到西方，还为我们创造了“zephirum”这

个词。该词可能译自阿拉伯语的“sifr”，它指代一个特殊的数字，以直立的蛋形符号表示。这个数字就是0。

源于印度的这套灵活自如的数字系统早就应该传播到西方世界了。662年，一位名叫塞维鲁·塞博赫特的叙利亚主教指出，“印度人”在天文学领域有了“一些微妙的发现”。他还特别强调说：“他们的计算方法极有价值，他们的计算能力高超到了难以形容的地步。更令人难以置信的是，他们在计算时竟然只使用了9个符号。”给塞博赫特留下深刻印象的是不包括0在内的其余9个数字，从这里可以看出古希腊学者的眼光是多么狭隘。他们似乎认为，希腊以外都是蛮夷之地。不幸的是，正是这种狭隘性导致他们对这套系统视而不见。10世纪末，法国数学家热尔贝（也就是后来的教皇西尔维斯特二世）曾经试图推广不包括0的阿拉伯—印度数字系统，但是他的努力没有成功。

尽管我们无法确切知道现代数学中真正的0起源于何时，但是我们至少可以推测阿拉伯数学家是从印度数学家极为先进的研究成果中汲取这个概念的。根据数学家、科学作家阿米尔·阿克塞尔在《零的起源》（*Finding Zero*）一书中的描述，他曾经花费大量时间，试图寻找在阿拉伯人之前使用0的相关记录。他希望可以找到确凿的证据，证明0既不是欧洲人的发明，也不是阿拉伯人的首创。他指的是20世纪20年代法裔匈牙利学者乔治·克代斯在研究活动中翻译的一篇柬埔寨语碑文。从寺庙残骸推断，树立这块石碑的日期是塞种纪元605年。我们知道塞种纪元始于公元78年，因此这篇碑文的起始日期应该是公元683年，这是人类使用0的最早的明确记录。

这篇碑文的重要意义在于0（用一个点表示）作为占位符出现在605这个数字之中。这是人类使用0的早期记录，其中还记载有日期（但它与古巴比伦人很早之前使用的占位符“\\”并没有本质上的不同，还不能证明人们已经开始使用真正的0了）。1931年的《伦敦大学亚非学院院刊》记载了这个发现，但是那篇碑文却已经遗失，而且没有留

下任何照片。因此，它的真实性就完全取决于克代斯报告的真实性。后来，苏门答腊出现过历史几乎同样悠久的记录，记载了人类在公元684年使用0的情况，但是阿克塞尔认为那篇碑文才是最早的已知记录，因此他决定进行追踪研究。

经过漫长的搜索，阿克塞尔终于在柬埔寨暹粒发现了“K–127”号石碑。碑文很完整，那个神秘的数字605也赫然在目。在外行人看来，那个代表0的圆点只不过是一个普通的小洞，更像后期风化造成的瑕疵，但是在内行人看来，这是人类早期使用0的证据。然而，热情高涨的阿克塞尔却发现，很难证明这个圆点与古巴比伦人使用的占位符有多大区别，具有全部数字功能的0似乎是沿着另一个方向进化而来的。

大量证据表明，早期的印度数学和天文学受到了古希腊人的影响，托勒密等人用作占位符的圆似乎就是从希腊传到印度的，但是，让0开始履行数字的全部职能的那次突破，却是印度当时的数学家促成的——正是他们无与伦比的创造性思维推动0发生了脱胎换骨的变化。最晚从6世纪开始，印度数学家就在使用“无”的概念，也就是没有任何数量的意思。但是，我们却无法确定，0作为一个数字到底是何时出现的。令人困惑不解的是，这个符号在当时身兼二职，既可以用作表示空缺的占位符，也可以用来表示未知量（就像代数老师告诉我们的那样，“用x表示未知数”）。

由此可见，占位符与未知数被混为一谈，原因是两者都代表一个空位。我们甚至可以从表示0的单词——“null”（或“nil”）看出两者之间的这种联系。这个英语单词来自法语的“null”，但它最初来源于拉丁文短语“nulla figura”，意思是“没有数字”。从古希腊数学到阿拉伯数学，其间有很多次数字0都呼之欲出，但是，印度人把0作为占位符使用的最古老的确凿证据却记载在一块瓜廖尔出土的876年的石碑上，上面的两个数字——270和50中的0都用一个小圆圈表示。然而，人类用圆圈表示0的做法很可能比这个时间更早。

7世纪，伟大的印度数学家婆罗摩笈多开始用0表示一个数减去自身之后的得数，但是，这种用法在此之前肯定已经非常成熟了。因此，我们现在使用的0的所有职能很可能有两个来源：一个是通过古希腊人传承下来的古巴比伦人的思想；另一个是后来在印度得到了更广泛应用的来自远东的数学理念，之后经由印度进入了正在兴起的阿拉伯科学和数学领域，这为增强数字系统的复杂性提供了一个全新的舞台。0是通过后一条路径最终进入欧洲的（随之而来的还有印度数字系统）。正因为如此，我们现在仍然把我们使用的数字称作“阿拉伯数字”，而不是更准确的“印度数字”。

正统的数学史肯定需要认真研究中国、南美洲和中美洲的数学发展情况，尤其不能忽视早期印度数学家的研究成果。但我们关注的是数学与科学之间的关系，数字0的使用以及印度数字对发展现代科学所做出的更大的贡献。三角学的正弦函数可能是印度数学做出的第二大贡献，这个概念值得我们关注，因为它推动印度数学在摆脱对现实的依赖这个方面达到了当时全世界的领先水平。

三角学的意思就是三角形测量，它的研究对象是三角形中的角与直线。在这个研究领域，古希腊人使用的是一种名叫弦表的复杂系统，而印度人则引入了正弦这个现代概念。所谓正弦，就是直角三角形中某个锐角的对边与斜边的比。大家在学校里应该学习过这个概念，但是很可能已经忘记它的含义了。从某种意义上讲，这个概念代表了印度数学在抽象化方面取得的突出成就。三角形的边和角都是清晰可见的存在，但正弦是一个比值，如果你不知道它的来历，就无法确定它的含义。正弦是一个非常有用的概念，但是与它的组成相比，毫无疑问它算不上一种真实的存在。

与正弦相比，0是一个更大也更重要的概念。但是，当斐波那奇将这个概念引入西方世界时，人们的反应却是褒贬不一。数学家们的热情程度似乎高于普通人的第一反应，他们迅速掌握了0的用途。17世纪

20年代，诗人约翰·邓恩在一篇布道文中抱怨说：“任何事物，数量越少，我们就越不了解。因此，0这个东西是多么难以看清、难以捉摸啊！”仔细琢磨这句话，数学家的热情与普通人的冷淡显然就很好理解了。

对这个全新数字系统的消极态度并不完全是情感造成的。会计人员发现，这个新系统中的0可以非常方便地被篡改成6或者9，因此不可避免地产生欺诈行为。为此，1299年，佛罗伦萨市议会颁布了一项法令，规定记账时必须用文字表示各种数目，而不得使用阿拉伯数字，以防遭到篡改。即使到了伽利略生活的时代，一位比利时牧师在与供货商签订合同时，还警告他们只能使用文字来表示所有数字。

但是，在这套新的符号与功能强大的0推动欧洲数学取得蓬勃发展之前，中东的另外一群数学家就已经接受了这套印度符号系统，并将它传播开去。我们知道，这套符号系统由中东传播至西方世界，因此被称为阿拉伯数字。但是，令数学研究发生翻天覆地变化的却是波斯作家花剌子米（al-Khwarizmi）(注)在825年前后创作的一部重要著作——《论印度数字的计算》（*On the Calculation with Hindu Numerals*），拉丁语版本的名称为“Algoritmi de numero Indorum”。我们把计算机执行的一系列规则称作“算法”（algolrithm），这个词就是从花剌子米的拉丁名字演变而来的。“代数学”（algebra）这个词也是由花剌子米创造的。有了这套灵活方便的数字系统之后，数学为我们创造了一个可以与丰富多彩的物质世界相媲美的全新世界。

代数学可能令很多学生头疼不已，但是它解决难题的能力和开放式方法的效力无与伦比。回头看看古希腊人研究数学的方法，将有助于我们了解印度数字和代数学等概念所具有的强大威力。古希腊人之所以一离开简洁美观的几何学就寸步难行，不仅因为他们处理分数的方法受到诸多限制，应用时困难重重，还因为他们没有办法处理我们

现代人借助代数就可以轻松解决的问题。一旦离开几何图形和直观思维，他们面对数学问题时就会束手无策。

以A + B = C + D这个简单的等式为例，古希腊人不会用符号的方式来处理这个步骤，而只能用文字来表述。而且，由于当时的习惯是单词之间不留空格，因此，上述等式的古希腊表达可能是：

THEAANDTHEBTAKENTOGETHERAREEQUALTOTHECANDTHEDTAKENTOGETHER（A与B的和等于C与D的和）

真实的情况可能会更加糟糕，因为古希腊人没有用字母或其他简单符号表示未知量的传统（如果他们真的使用字母，就会特别难以理解，因为他们在表示数字时也会使用字母）。也就是说，“文字等式”中不会出现A、B、C和D，而是用文字把需要加到一起的事物一一表述出来。这个例子再一次说明，古希腊人放弃古巴比伦人的方法是一个退步。古巴比伦人对数字的态度有所不同，他们可以解决代数学难题，甚至可以解某些二次方程。然而，二次方程的这种解法被遗忘了有千年之久。

然而，这不能阻挡希腊数学前进的步伐。从公元250年前后至亚历山大里亚学派[注]后期，代数学研究已经得到了中等程度的发展，其中的典型代表人物是丢番图[注]。丢番图似乎重新拾起了古巴比伦人更重视数字的研究成果，还加入了古希腊研究的精髓——精确性和证明，忽略了近似表示。丢番图的最大贡献或许是他创立了符号法，代数问题从此变得简明扼要，难度也大大降低了。

在《算术》（*Arithmetica*）一书中，丢番图利用符号和结构来表示未知数、10的幂和算术运算。他给出的方程与我们现在使用的方程并不完全相同，而是用一种简单明了、前后一致的符号表示。以$2x^4$ +

$3x^3 - 4x^2 + 5x - 6$为例。如果用字母S表示平方，C表示3次方，x表示未知数，M表示减，u表示数字1，丢番图的方程就是SS2 C3 x5 M S4 u6。他的这些贡献，成为代数学蓬勃发展的基础。

花剌子米的一本关于代数学的著作（*Al-jabr wa'l muqābalah*，书名意思不明）仅使用文字表述，甚至都没有使用丢番图发明的笨拙的方程。因此，从某种意义上说，这部著作相对于丢番图的某些研究来说是一种倒退。但是，花剌子米的著作与现代基础代数入门读物很相似，因为他给出了方程的解法，尤其是二次方程，尽管是以文字的形式表述。有人认为，阿拉伯世界当时实施的遗产继承规则十分复杂，然而，遗产金额的计算却意外推动了代数学的发展。花剌子米在解方程时通常会忽略负根，可能也是出于这个原因。

随着花剌子米等阿拉伯人的著作被翻译流传开来，情况开始发生变化。我们知道，在斐波那奇出版《计算之书》后的几百年时间里，这套新的数字系统一直未被普通人接受。但是，与其说这是一种明智的行为，不如说这是旧有系统垂死之前的最后挣扎。那些善于使用算盘或计数表（与算盘比较相似，但是不穿绳）的既得利益者持抵制态度，可能是担心引入这些新数字后，他们在这方面的特长就会一文不值。从这个意义上看，斐波那奇选用的书名的确有几分讽刺意味。毕竟，我们都曾抱怨"新数学"与我们在学校里学到的数学并不是一回事。但是，变化是无法抗拒的。新系统的优势是非常明显的，所有对数学有所研究的人都知道，新系统被人们接受几乎是大势所趋。

在中世纪学者中，有很多人故意贬低数学的重要性，但有一位学者发起了一场运动，为人们认识数学的重要性扫清了障碍。

---

1. 婆罗摩笈多（598—约655），印度天文学家、数学家，著有《婆罗摩修正体系》。——译者注

2. 婆什迦罗（1114—约1185），印度数学家、天文学家，著有《历算书》。——译者注

3. 花剌子米（约780—约850），数学家、天文学家，被誉为“代数学之父”。——译者注

4. 亚力山大里亚学派是古希腊数学家在埃及亚历山大城建立的学派，分前期（前4世纪—前146）和后期（前146—公元641）。前期以欧几里得、阿基米德等人为代表，后期以托勒密、丢番图等人为代表。——译者注

5. 丢番图（约246—330），古希腊亚历山大里亚学派后期的重要学者和数学家，代数的创始人之一，对算术理论有深入研究。——译者注

第7章
培根：数学是自然科学的钥匙

Are Numbers Real?

一提到“培根”这个名字，大多数人肯定会想到那位创立科研方法雏形的16世纪政治家、哲学家——弗兰西斯·培根。但是，弗兰西斯对数学的态度近乎漠不关心。在他眼中，科学研究不过是收集、整理、分类。后来，伟大的物理学家欧内斯特·卢瑟福嘲笑说，在培根眼中，“科学研究，除了物理，就是集邮”。我下面介绍的“培根”是罗吉尔·培根，与弗兰西斯·培根没有任何关系，但是他看待世间万物的观点似乎更具现代性。罗吉尔说：“对数学一无所知的人，不可能了解其他科学和世间万物。”当时的社会根本没有认识到数字和数学对于人类理解自然的重要意义，培根希望改变这种局面。

培根这个人似乎非常有趣，但是他出生于800多年前，因此关于他的任何信息都难以确认。除了流传下来的一些逸闻以外，我们掌握的主要资料就是他本人的作品，所以我们只能根据其中的蛛丝马迹来推测他的生平。通常，人们认为培根出生于1214年。

1267年，培根在他的代表作《大著作》（*Opus Majus*）中写道：“我一直在努力地钻研科学、研究语言，从我学习第一个字母以来，已经过去40年了……40年来，我潜心研究，心无旁骛，其间只中

断了两年。”人们通常认为，培根所说的40年的研究生涯（他在书中使用的拉丁语表达是“in studio”），是从他参加牛津大学入学考试的时间算起的，因为通过入学考试代表他正式进入了一所大学。当时，人们通常会在13岁时参加大学入学考试，据此可以推断出培根出生于1214年。但是，这就意味着培根说的“学习第一个字母”是一个非常隐晦的比喻，其实是指他开始从事正规研究的时间。如果这句话真的是指他开始学习字母的时间，那么他的出生日期就要向后推，估计在1220年。但是，人们普遍认为他出生于1214年。

据15世纪的历史学家约翰·劳斯称，培根出生于伊尔切斯特这座寂静的乡村小镇。但我们没有找到有关培根出生地点的其他资料（培根本人也从未提过他的家乡）。我们只知道培根进入牛津大学学习，而且他到了这里之后做的第一件事可能就是理发。学生在上大学之前必须完成某些宗教仪式，成为事实上的小僧侣。其中一个必不可少的环节就是理发，要把头顶上的一大片头发剃光。牛津大学成立之后，校园里最早出现的生意之一就是理发店。牛津大学现有的学院在那时候都还没有成立，也没有现代意义上的大学建筑。学生们都住在出租屋中，至于上课的地方，则是在老师们在市区租用的大房间里。

就这一点而言，这所大学与后来那所家喻户晓的现代大学几乎没有任何相似之处。牛津大学的历史可以追溯至1095年，当时，来自埃唐普的教师西奥博尔德决定在牛津创建一所高等教育学校。这座城市并不是英国创建第一所大学最理想的地址。英国有很多城市都建有大教堂，有建造教会建筑的传统，而且修道院现成的学习中心可以为教学提供便利。但是，在12世纪斯蒂芬与玛蒂尔达之间的战争爆发之后，烧遍全英国的战火证明牛津大学的选址是具有战略眼光的。英国人在为第一所大学选址时，选中了这里，也许就是因为牛津位于英国的中心位置，交通便利。

1209年，就在学校当局准备效仿巴黎大学（成立于12世纪中叶）大干一番的时候，一桩谋杀案差一点儿让牛津大学毁于一旦。由于某名学生（所有学生都应是禁欲者）的情人被谋杀，愤怒的市民仿佛变成了嗜血暴徒，叫嚣着要学生们血债血偿。政府官员立刻做出判决，绞死了另外两名学生，只是因为他们在错误的时间出现在了错误的地点。

学校当局没有弄明白一个问题：市民逼迫绞死无辜学生，与市民插手学校内部事务，这两种情况到底哪一种更难以忍受？结果，有70名老师（在全体教职人员中占相当高的比例）离开牛津市，来到东英吉利亚地区的一个名叫剑桥的落后小镇，加入了不久前创立的几所学校。庆幸的是，教会特使尼古拉斯来到英格兰规劝桀骜不驯的国王约翰，使牛津市恢复了秩序。1214年，牛津大学的基本结构被确定下来，但是直到1231年，才获得正式的特许证。

罗吉尔·培根入学时（可能是在1227年），牛津大学还是一所历史非常短的新学校。尽管从理论上讲，这所大学属于教堂体系，但是校园里洋溢着狂躁，甚至是危险的气息。贫民区中的致命攻击时有发生，有的是暴力殴打造成的，有的则是因为市民与大学生之间发生冲突。这些大学生的着装与城里的年轻人区别不大，但是他们颇有特点的发型肯定会暴露他们的身份。此外，来自英国南方和北方的学生因为分属不同的派系，也会发生激烈的打斗。

根据1238年的一份报告，我们可以很好地了解培根在牛津求学时所处的环境。这份报告称，新任教会特使奥托即将造访牛津市，并将在西南方不远的奥斯尼修道院下榻。一群老师和学生来到修道院欢迎特使，但是在这个愉快的欢迎仪式上却出了大乱子。一位爱尔兰学生在修道院门口祈祷时，修道院的大厨将一锅开水倒在这位不幸的学生身上。然后，另外一名学生掏出弓箭，射杀了这位厨师（即使在社交场合，携带适当的自卫武器也是必要的，是可以接受的）。

欢迎的人群一下子变成了群情激奋的暴徒。为安全起见，教会特使被偷偷地护送到了泰晤士河下游的沃灵福德。整个大学被严密控制，几个月后，学校中担任管理职务的老师们光着双脚，沿着伦敦大街走到教会特使住所请罪，之后情况才有所好转。培根说话一向直言不讳，又没有从事这种政治活动的天赋，要在学校里出人头地非常困难，因此他不大可能从事管理工作，应该也没有参加这次活动。至于那些学生，他们很快又恢复了以往的狂暴作风。

攻读学士学位大概花了培根6年时间（别忘了，中世纪大学教育的前半部分基本上是传授中学教育的内容），攻读文学硕士学位又花了他两年时间，因为担任大学教师必须拥有文学硕士学位。接下来，他可能又花了8年时间获得神学硕士学位。要获得只有神学领域才有的博士学位，他还需要再花8年时间潜心研究，但我们没有找到相关证据。之后，培根来到巴黎大学。由于法国曾经禁止各大学介绍亚里士多德的成果，当他们重新将这些内容纳入教学范围之后，却发现在本地已找不到可以胜任的老师，所以牛津大学毕业的硕士在巴黎大学深受欢迎。当时，西方国家普遍认为巴黎大学是最好的三所大学之一，另外两所是牛津大学和剑桥大学。

培根在巴黎大学只待了几年，但是他在那里结识了马里库尔的彼得。彼得是一个非常神秘的家伙，人们称他“流浪者彼得”（Peter Peregrinus）。现存于世的最早研究磁力的论文中，有几篇就是彼得撰写的。彼得将培根的兴趣引向科学，在他的启迪下，培根对实验产生了极大的热情。13世纪40年代后期，培根回到牛津大学，花了很多钱购买书籍和实验工具。1267年，培根写道：“20年来，我放弃了寻常的方法，专门研究智慧。我花了2 000多英镑，购买各种书籍、语言学习材料、器材和数学用表等。”当时，建造一幢大房子的成本是2—3英镑。尽管培根的这个说法有些夸张，但是为了科研事业他肯定动用了家庭财产。数学已经成为培根理解周围世界的一个重要工具。

回到英格兰之后，培根很快就加入了一个成立不久的宗教团体——方济各会，原因可能是他的钱花完了。这些穿灰袍的方济各会修士正在牛津市建立一个庞大的机构，覆盖的范围从牛津市南门一直延伸到西边的城堡。与圣本笃会等成立时间更久的修道会的修士相比，方济各会的修士享有更多的自由。而且，在牛津，宗教团体还与大学建立了紧密的联系。加入方济各会后，培根得以接触到大量书籍，而且有充分的时间研究自然哲学。自此之后，培根开始了无忧无虑的生活，直到1250年，他被派遣到巴黎的女修道院。培根说，他在那里待了10年，做的都是“低贱的工作”。这些遭遇很可能是因为培根经常直言不讳地表达自己的观点，令权威人士深感不安。

在巴黎逗留期间，培根产生了改革历法的念头。他发现当时使用的儒略历还是罗马时代的产物，一年的时长比精确值大约多出11分钟。计算结果表明，每125—130年（确切数字是128年），该历法给出的四季交替时间与实际情况就会相差一天。差距似乎并不大，但是从沿用这套历法以来，总的时差已经超过一个星期了。这意味着宗教节日的日期是错误的，这让培根无法容忍，因此他提出改革历法，但是他的申请石沉大海。直到1582年，基督教国家才开始采用格里历，这套历法与培根当年详细描述的方法几乎一模一样。而英国直到1752年才接受新的历法；美国受当年殖民统治的影响，全国各地大多采用各自殖民者的历法，这种混乱局面持续了几十年时间。

培根在巴黎的时候，方济各会为了重新扎根于穷困潦倒、缺衣少食的人群，通过选举产生了新的领导人。这位领导人决定引领这些灰衣修士远离学术机构，还禁止他们著书立说。但是，培根难以割舍那份对科学的情怀，因此他多方努力，希望可以绕开这条禁令。他写信向几位重要人物求助，其中包括教会派驻英国的特使、主教居伊·德富尔克。培根告诉主教，自己想写一本科学方面的书，但又不敢违背禁令，因此恳请得到主教的特许。但是，德富尔克误解了他的意思，在回信中说他希望马上看到这本（还没动笔的）书。

培根写那封信的本意其实是寻求资助。由于家里的钱已经用完了，方济各会又不提供支持，他需要从外界得到资助，才能着手收集资料。结果，德富尔克不但没有提供资助，反而提出了一个要求。就在培根焦虑不安，不知道如何回应主教的这个要求时，他收到了一个意想不到的消息。1264年，德富尔克被召唤到佩鲁贾。原来，枢机团注选举他担任教皇。第二年，他加冕成为教皇克雷芒四世。一夜之间，培根的这位身居高位的朋友登上了权力的巅峰。

1266年，克雷芒四世写了封信，允许培根着手进行科学研究，并授权他可以不遵从禁令。但是，培根仍然需要想办法筹集资金，他试图向朋友求助，但是朋友的支持非常有限。1267年1月，他决定向教皇递交一份简短的写作大纲，介绍他未来的那部代表作——《大著作》，恳请教皇提供资金支持。写过非小说类作品的人都知道，撰写创作大纲是一项非常痛苦的工作，需要作者将整本书的内容压缩成简明扼要的介绍。

培根无法满足这个必要的限制条件，他的“简短”大纲最终变成了一本厚厚的书，篇幅超过50万个单词（大概是本书的6倍），内容涉及光学、天文学、力学、炼金术、农业和医药业，最后还用了一章着重讨论实验科学。在请人誊写这份大纲时（当时，创作全部是用纸笔完成的），培根想，这份大纲太长了，已经失去了大纲的意义，因此他需要再附上一封信。结果，这封信也写得很长，变成了另外一本书。令人惊奇的是，这一幕第三次上演，第三本书就这样产生了。最后，他寄给教皇的是三本书，共计约100万个单词，创作时间总计12个月。（实际上，他当时只寄出去两本书，因为信使出发时，第三本还在誊抄过程中。）

这时候，培根已经开始整合牛津大学的资源，准备撰写他那部酝酿已久的大作。但是，从欧洲大陆传来的消息令他万念俱灰——克雷芒四世死了。而此时，他寄给克雷芒四世的信还在路上。在教会中，

有人认为只需要掌握神学知识就可以了，有人对科学则一无所知，所以教会中几乎没有人支持培根。于是，培根对这种现象发起了猛烈的抨击。写于1370年的一部编年史称，培根因“思想怪异”而受到谴责。他可能被关进了意大利中部城市安科纳的一所监狱，直到1290年才被释放。根据劳斯的记录，培根死于1292年。

培根其实不算一名伟大的数学家，但是他留下来的《大著作》一书却是一部无与伦比的杰作，详细地描述了当时已揭示的自然奥秘。这本书最特别的关注点是，数学对于自然哲学家（甚至神学研究者）的重要意义。培根在学术上的前辈罗伯特·格罗斯泰特（Robert Grosseteste）也非常重视对数学本质的研究，他们两个人显然为未来的研究指明了方向。此外，培根还指明了历法改革的发展方向。正因为如此，他把自己放到了一个非常有趣的所谓“祖父效应”的位置上。

我们习惯于给某个人贴上“某某之父”的标签，这里的“某某”是指处理某个问题的全新方法，例如“计算之父”、“太空飞行之父”。但我认为，“祖父”的角色有些微妙的区别，“祖父”会指出一个非常宽泛的方向，或者一条充满艰辛的道路，但是通常不会告诉人们如何沿着这个方向顺利地走下去。一段时间之后，“父亲”这个角色就会出现在继承结构的直线上，并且朝着最终结果迈出非常重要的一步。

我认为，维多利亚时期的摄影师埃德沃德·迈布里奇就是电影的“祖父”。（他的真名叫作爱德华·马格里奇，但是他行事高调，经常重塑自己的形象。）他先是为加利福尼亚的铁路大亨利兰·斯坦福工作，后来进入宾夕法尼亚大学。他通过连续触发一组照相机的方式，拍摄出马、人和其他动物运动的系列照片。利用现有条件拍摄出这些静止照片，本身就具有重要的价值，可以提供大量信息。迈布里奇还借助一台简陋的投影仪，将这些照片转变成真正的动态影片，尽管影片的长度非常短。1893年，他甚至还在芝加哥郊区举办的哥伦比亚世界博览会上建造了第一个专门放映电影的剧院。

迈布里奇出人意料地把这座影院命名为“动物实验摄影厅”（Zoopraxographical Hall）。他在这座全尺寸专用建筑中举办教育讲座，介绍自己的成果。但是实际上，观众们花25美分购票的目的是观看大屏幕上的拳击比赛和衣着暴露、来回走动的美女。（迈布里奇喜欢捕捉衣着暴露或者赤身裸体的人物形象，他甚至拍摄过表现裸体泥瓦匠干活场景的系列照片，目的是研究他们的肌肉组织。）尽管迈布里奇的照片质量得到了人们的一致认可，但是在相当长的时间里，他的动态影片却遭到无视，部分原因是一位电影史学家发起了一场联合抵制运动，号召人们忘记他的这些作品。当然，迈布里奇的技术注定不会成功，因为主流技术需要使用胶卷。但毋庸置疑的是，他展示了电影的概念，并因此赢得了“祖父”的地位。

同迈布里奇一样，培根并没有昂首阔步走在数学领域的发展道路上，但是在大多数学者对数学嗤之以鼻的时候，他却独树一帜，强调了这门学科的重要性。当时，神学是唯一可授予博士学位的学科，因此培根很可能有一种自卑情结。但是，他仍然指出，当代的神学研究者对数学的重要性不屑一顾，是因为他们不知道数学可以发挥什么样的作用。现在看来，造成这个局面的部分原因是数学与魔术之间的联系让人们感到不安。

培根本人对于魔术是持严厉批评态度的，他指出这是骗子常用的伎俩，目的是从那些没有受过教育的人身上骗取钱财。但是在当时，很多人，甚至包括某些学者，都会把魔术与数学混为一谈。这个误会是一个非常不幸的巧合造成的。表示科学知识的单词读作“matesis”，而表示占卜的单词读作“mathesis”，但是这两个单词都写作“mathesis”。甚至到了都铎王朝，还发生过类似的现象，人们用“calculating”（计算）这个词表示玩魔术，将数学文本与介绍魔术的文本混为一谈的情况时有发生。

毫无疑问，罗吉尔·培根在数学上的造诣并不深。他在《大著作》一书中就犯了一个低级错误。为了证明亚里士多德的知识不全面，他举例说这位古希腊哲学家曾承认不知道如何化圆为方，接着他又说，“如今，人们已经非常清楚如何解决这个问题了”。但是，我们知道，化圆为方，或者说只使用几何学的常规工具——直尺和圆规，画出一个与圆面积相同的正方形，自古以来就是一个令数学家头疼不已的问题，而且人们已经证明这个问题是无法解决的。

尽管培根水平有限，但是他敢于在那个时代强调数学的重要性，毋庸置疑是非常了不起的。17世纪末，与艾萨克·牛顿同时代的英国数学家约翰·沃利斯给牛顿在数学领域的冤家对头戈特弗里德·莱布尼茨写了一封信：

> 在本世纪，一些人（追随伽利略的步伐）把数学归入自然哲学的范畴，从而大幅度推动了物理学的发展。400年前（甚至更早），罗吉尔·培根（从黑暗的百年历史中脱颖而出的一位伟人）也做过同样的努力。

在《大著作》一书中，数学整整占据了一章的篇幅，尽管其中还涉及历法改革、占星术等其他内容。对于那些从小就被告知占星术根本不科学的人而言，把占星术纳入其中可能是一个令人扫兴的选择，但是培根毕竟要受到所处时代的影响。培根对占卜算命者所用的占星术嗤之以鼻，认为占星术就是在利用轻信心理骗取普通人的钱财。但他也认为，出生时天体的排列分布有可能影响人们的性格。正因为占星术的这种表现形式（一种先天与后天之争），培根那个时代的人认为占星术有可能发展成一门潜力巨大、地位崇高的科学。

天体的排列分布几乎不可能对人产生影响，但是培根那个时代的人无从了解这个事实。考虑到当时的科学水平，占星术有可能影响人物性格这个说法（与某些现代企业钟爱的迈尔斯—布里格斯人格类型

测验有几分相似），在那些从事这项活动的人看来是有道理的。培根认为，科学在很多方面都要借助数学这个工具，原因是科学可以从数学中获益：

> 数学是自然科学的大门和钥匙……对数学一无所知的人，不可能了解其他科学和世间万物……更糟糕的是，对数学一无所知的人不会认识到自己的无知，因此也不会考虑如何做出补救。而且，掌握这门科学之后，我们的大脑会更加聪明，为清楚了解世间万物做好准备。因此，如果大脑理解了这门科学的基础知识，并且利用这些基础知识来探查其他科学以及世间万物，所有问题都将迎刃而解，不会留下任何疑问，也不会造成任何错误。

对于培根而言，数学不仅是一种工具、一种思维方式，还是逻辑的一种结构化应用。不懂数学的人无法理解，它可以帮助大脑更好地思考并理解自然。接着，培根论证了数学对于我们理解其他科学的重要意义。以光学（培根深入研究了这门科学，不仅提出了一些新颖的理论，还完成了一些实验）为例，要学好这门科学，首先必须正确理解角、几何学和对称等相关内容。

认为应该将数学纳入教学范围的不止培根一人。（当时的教育已经包括了一些数学内容。由算术、几何学、天文学和音乐等4门学科组成的“四艺”在大学课程设置中占有相当大的比例，这4门课要么直接属于数学，要么核心内容与数学密切相关。）但是，培根看到了数学对于各门科学的重要意义，他将数学从自己的独立王国中解放出来，指出它可以为我们了解自然界提供帮助。这才是培根出类拔萃的地方。

总的来说，培根的观点在几百年时间里都没有引起人们的注意。然而，数学王国中也曾有几道闪电划过黑暗的天空。其中一道“闪电”来自牛津大学，即“牛津计算师”，也称作默顿学派、默顿数学家。他们人数不多，但都与牛津大学的默顿学院（牛津大学最早成立的学

院之一）有关，他们充分认识到了数学（和推动数学发展）的重要性。

这个学派（托马斯·布拉德沃丁和威廉·赫特斯柏立是其中的代表人物）研究逻辑、几何学和计算方法，但是他们最著名的成果是落体定律，即“平均速度定理”。该定理称，在相同的时间内，如果第一个物体做匀加速运动（例如由高处掉落），第二个物体以第一个物体最终速度的一半做匀速运动，那么在该段时间内两个物体的运动距离相等。有意思的是，布拉德沃丁及其同事的研究之所以具有重要意义，最主要的原因可能在于他们的研究背离了亚里士多德对运动的理解，反而有点儿接近于伽利略和牛顿的观点。重要的是，亚里士多德研究数学的动力来自纯粹的哲学，而在默顿学派的眼中，数学可以帮助我们得出答案这个作用更重要。

14世纪的法国学者尼古拉·奥雷姆是拓展数学思维的又一个杰出代表。我们现在使用的幂运算就是他的研究成果。两个底数相同的指数幂相乘，直接把指数相加即可，例如，$y^2 \times y^3 = y^5$。此外，他还探索了分数指数幂（例如$x^{1/2}$）的可能性，但我们现在使用的幂的概念在当时还没有确定下来，因此他使用的是一种间接的研究方法。与默顿学派的成果相比，他在摆脱现实的抽象化道路上走得更远。默顿学派研究的正方形与立方体都与观察运动物体的特性有关。

奥雷姆可能是第一个提出用图像表示数学结构的人。这个成果不是很抽象，但是重要性却丝毫不减。今天，我们理所当然地认为图像表示法非常有效，不仅可以帮助我们更直观地了解函数（例如函数$y = x^2$），还可以帮助我们深入研究微分和积分等概念（在介绍微积分时，我们将讨论这两个概念）。奥雷姆至少是推广这个方法的第一人（当时称作“形态的幅度”），他甚至还考虑了将这个方法推广至三维空间的可能性。

通过强调数学和实验的重要性，培根及其中世纪的追随者们在当时的自然哲学与未来的真正科学之间架起了一座桥梁。对于培根来说，数学是一种工具，可以帮助我们理解世间万物，但是它本身绝不能真正地摆脱与现实的联系。他从未接触过虚数和复数的概念，也许反而是件好事。

---

1. 枢机团或称枢机院，是天主教会的最高宗教机构。——译者注

第8章
高斯：神通广大的虚数

Are Numbers Real?

数学家把负数的平方根称为虚数。乍看上去，这个概念似乎非常奇怪，完全背离了在这之前数字与现实世界建立起来的关系，仿佛是要证明数学与现实世界彻底断绝了联系。然而，随着时间的推移，人们在探索物质世界的过程中却发现虚数的作用异常广泛，而且关注现实的工程技术人员也离不开虚数，在每天的计算中都要用到它。

说到虚数，首先要从负数说起。数学家发现自然数的这些变体可以构成一个实用的数学概念之后，就对它们的数学特性进行了考量。他们发现，两个负数相乘就会得到正数。其中的道理并非一目了然，但是借助数轴还是很容易看清楚的。减号实际上表示数轴上的方向变化，因此，在正方向上变化两次方向，朝向的仍然是正方向。我们也可以换个方法，认为这是数学家做出的一个主观决定，但他们也是不得已而为之。对不同的决定稍加研究就会发现，这是与数学其他方面保持一致的唯一选择。

接下来，新的问题出现了。既然我们已经知道正数乘以正数的结果是正数，负数与负数相乘的结果也是正数，那么负数的平方根是什

么呢？什么数的平方是负数呢？答案既不可能是正数，也不可能是负数。人们发现，可供选择的答案似乎并不多。

我们暂且不要急于沿着这个思路继续考虑下去，而是回过头来思考负数与现实的关系。我们从第2章得知，负数在记账的时候是有用的，它可以表示债务。在邻居来借山羊时，我们可以用负数表示与原来相比山羊减少的数量，也就是邻居借走的山羊的数量。但是，在这些例子中，其实仍然用的是正数，只不过形式特别。比如，我没有办法表示我欠你多少钱，我只能告诉你我需要还给你多少钱。那么，我能在现实世界中找出明确表示负数概念的东西吗？事实上，真的可以找到，条件是允许我突破自然数的范围。但是，在19世纪之前，没有人认识到这种可能性。

大家想一想电池两极上的标志：一个是正号，一个是负号。我们通常认为这种标记方法是本杰明·富兰克林发明的。实际上，刚开始的时候，这两个符号仅为了表示两者的不同，而不代表数学中正数与负数之间的区别。然而，我们现在都知道电子与质子携带的电量相同，电性相反。它们没有方向，本质上是纯粹的数值（数学家和物理学家称之为“标量”），与牛顿第三定律描述的大小相等、方向相反的作用力不一样。但是，它们与正负数一样，也可以相互叠加或者抵消。

虽然正负电荷的定义具有主观随意性，但是这些电荷都是实实在在的东西，特性与正负数相似。这似乎说明通过现代科学发现也能找到现实世界中的有理数。现在，我们知道质子、中子等粒子是由更小的粒子——夸克构成的，而夸克携带的电量是基本电荷的2/3或–1/3。尽管这些数值都不是近似值，而是一丝不差的精确值，而且看上去都是有理数，但是我们在将质子电量定义为1，将电子电量定义为–1时，我们并不知道这些粒子的电量到底是多少。实际上，夸克的电量是2或者–1，而质子和电子的电量分别是3和–3。

我们接着讨论负数的平方根。在数学家发现他们无法利用已有的数字表示负数概念时，他们就很随意地做出了一个决定——创造一种新的数字。他们之所以这样做，并不是因为他们需要负数，而是因为他们勇于探索，渴望了解数字世界的未知领域。笛卡儿不无讽刺地为这些数字赋予了一个非常妥帖的名字——“虚数”，这些稀奇古怪的数字也变成了数学家的新宠儿。数学世界仿佛又增加了一个维度，而且这个新维度似乎与物质世界不存在对应关系。刚开始时，虚数不过是数学家的玩具，但是这些玩具的灵活性高得惊人，适用于所有的数学运算法则。

米兰医生、数学家吉罗拉莫·卡尔达诺[注]在《伟大艺术：代数法则（第一册）》（*Artis Magnae Sive de Regulis Algebraicis Liber Unus*）中第一次提出虚数的基本概念，这本著作出版于16世纪上半叶。这本书之所以有名，原因可能与这本书在一定程度上涉嫌欺骗有关。卡尔达诺从同为数学家、工程师的尼科洛·塔尔塔里亚那里学会了三次方程（未知数的最高次数是3的方程，例如$x^3 + 4x^2 + 2x + 5 = 0$）的解法，并保证不会把这个方法告诉别人。但是，卡尔达诺把这个方法写到他的书中，并且公开发表了。他在致谢中充分肯定了塔尔塔里亚的贡献，因此这不算是剽窃行为，但是他显然没有遵守自己的诺言。

卡尔达诺还在书中顺便谈到了一个看起来无伤大雅的简单方程$x^2 + 1 = 0$，并讨论了它的解法。这个方程与$x^2 = -1$同解（第一个方程两边同时减去1就可以得到第二个方程）。除非某个数的平方是负数，否则这个方程无解。卡尔达诺说，这样的数“不仅没什么用处，而且难以捉摸”。后来经证明发现，这个评价与事实相差甚远。19世纪，德国数学家卡尔·弗里德里希·高斯发现虚数可以方便地拓展数轴，形成二维的数字平面，至此，虚数的价值才表现出来。

我们已经知道，人们在考虑整数时，经常会想象一条左右两端分别向负无穷和正无穷延伸的水平直线，整数排列在这条直线上，0位于

正中的位置。高斯沿着与这条直线垂直的方向画出了第二条直线，并让正虚数排列在朝上的方向，负虚数排列在朝下的方向。这样，平面上的任意点都可以用一个“复数”加以定义。复数是由一个实数和一个虚数构成的数。如果用i表示-1的平方根，那么$5 + 2i$就是一个复数，可以用来定义实数轴（横轴）上有5个单位、虚数轴（纵轴）上有2个单位的那个点。

我们也许会认为，这不过就是将坐标系上我们都非常熟悉的x轴、y轴换了个名称而已。但是，有了这个变化之后，我们就可以把复数当作普通数字进行代数运算，应用代数学的所有法则和方法，并最终得出适合这个二维空间的结果。事实证明，复数是描述各种波的理想选择，因为这些波天然地具有二维的形态。于是，在不知不觉间，虚数就无处不在了，从简单基本的电场计算到复杂深奥的量子力学方程，我们都可以看到它们的身影。只要在计算结束时舍弃虚数结果，就不会产生虚数值电流这样的结果。实践证明，复数的用途广泛，是一个强大的数学工具，并将继续发挥它的强大作用。

虚数是否真实存在呢？虚数显然不会与物质世界中的任何事物构成直接的对应关系。比如，我没有办法拿出$3i$个苹果。我可以从已有的苹果堆中拿出3个苹果，来表示-3个苹果，但即使是这种间接的办法，也无法表示$3i$个苹果的概念。然而，任何东西，只要它可以严格定义，而且遵从所有规则，就可以在数学世界中找到立足之地。由于虚数（特别是复数）可以有效地表现二维空间的变化，因此在它们历经艰辛闯入抽象的数学世界之后，人们发现它们竟然是解决现实世界难题的绝佳工具。

虚数提供的是一个现实世界所没有的工具。我们可以将现实世界的难题搬到虚数世界中，用虚数提供的特有办法加以解决，然后再把它们送回现实世界。自然数非常简单，我们可以认为它们与一个物体

或者一群物体存在直接的对应关系。虚数和复数则活跃于一个平行世界中，但它们仍然可以启发我们，帮助我们了解物质世界的奥秘。

然而，在虚数真正地在实践中大放异彩之前，人们仍将借助传统而直接的正数和几何学，去征服神秘的宇宙。

---

1. 吉罗拉莫·卡尔达诺（1501—1576），意大利文艺复兴时期的全能学者，古典概率论创始人。——译者注

第9章
牛顿：微积分与宇宙观

Are Numbers Real?

数学经历了很长时间才在科研活动中找到充分展示自己的舞台，原因之一就是人们的宇宙观中普遍充斥着一种超自然的神秘性。自然哲学家认为除月球运转轨道以外的所有事物都是完美的，是由一种与宇宙中其他事物都不相同的要素（所谓的第五元素）构成的，它们能够运转，是因为天使给了它们动力。在这种情况下，很难想象数学可以发挥任何作用。但是，伽利略打破了古希腊的宇宙模型，开创性地利用数学来预测抛射体和钟摆的运动轨迹。

艾萨克·牛顿站在这位巨人的肩膀上，绘制出宇宙力学图。借助伽利略的研究成果，只要掌握足够的数学知识、完美的数据和超强的计算能力，数学家就可以洞悉一切。牛顿是一名基督徒（尽管不是正统的基督徒），因此他不会公开宣扬，但是作为他的继承者、超级拥趸，18世纪的法国学者皮埃尔–西蒙·拉普拉斯[注]却没有任何迟疑。拉普拉斯是当时不多见的无神论者。（据说拿破仑曾经问拉普拉斯，上帝在他的哲学中处于什么位置？拉普拉斯回答说：“我不需要做那样的假设。”）牛顿认为，宇宙是一个异常复杂的机械装置，就像一座大钟，只要掌握足够的信息，拥有健全的智力，预测未来就完全有可能做到。他说：

> 假设有一位智者，他能理解所有驱动自然的力，以及形成各种力量的对应环境，并且能够分析这些数据，他就可以用一个公式来表示包括最大天体与最小原子在内的世间万物的运动情况。对于他来说，没有什么东西是不确定的。未来如同历史一样，在他面前一览无余。

牛顿永远也不会成为这样的智者，然而，当数学开始在科研活动中占据重要地位的时候，他通过计算，并借助一种新颖而神秘的数学方法——处理无穷小问题的流数术，提出了牛顿运动定律和万有引力定律。牛顿取得的成就非常重要，地位举足轻重。他的数学也许不能完美地预测未来，但是在预测作用力（尤其是神秘的万有引力）的大小及其效果时却展现出引人注目的力量。但是，为了不让他的读者感到害怕，抑或是为了让他的方法显得高深莫测，牛顿在创作他的代表作《自然哲学的数学原理》时，煞费苦心地把很多研究成果转化成传统的几何学。牛顿的世界就像一个钟表，表现出确定性和可预测性。

牛顿的成果可以转化成几何学形式，可能要归功于他的法国前辈、哲学家笛卡儿的研究成果。笛卡儿有两件事让人们难以忘记："我思故我在"的宣言，以及以他的名字命名的"笛卡儿坐标系"。但这只是冰山一角，他的研究包罗万象，不仅涉及光学理论，他还试图从科学的角度研究灵魂。

笛卡儿坐标系的意义远不只是用一组数字表示一个点这么简单（早在13世纪，培根就已经知道这个方法了）。笛卡儿还创立了在几何形状与等效代数方程之间来回转换的解析几何学。例如，利用（x，y）坐标值，可以画出方程（例如$y = x^2 + 2x + 3$）的图像。同样，自然界中有很多变化过程也可以绘制成图像，从而与代数方程对应起来，这样就可以大大降低预测结果的难度。

笛卡儿的这个方法使奥雷姆利用图像表示函数的想法变成现实，它将在牛顿的研究工作中发挥巨大的作用，使他的大多数代数研究成果隐藏在几何学的外衣之下。笛卡儿本人似乎也没有意识到这个概念到底有多大的影响力，他在《几何学》（*La Géométrie*）中介绍了这个概念，并且说这是一个非常简单的构建几何形状的方法。从本质上看，他认为这与牛顿将代数学问题转换为几何学问题的方法比较相似，而没有像现代人一样，看到它将空间问题转换为代数学问题的真正威力。但是，无论笛卡儿的目的是什么，他都为我们创造了一个将几何学问题转换为代数学问题的方法。几何学与我们周围世界的联系更直观，而代数学则给人一种更抽象的感觉，大多数的现代数学研究都采用了这种方法。

牛顿在数学领域取得的杰出成就——流数术古已有之，可以追溯至古希腊时代。古希腊人对图形不断进行分割，使之变成许多尽可能小的形状，然后计算这个图形面积的近似值。例如，如果想计算圆的面积，我们可以想象沿着半径的方向，将圆分割成一系列橘瓣状的平面图形。随着这些“橘瓣”越来越窄，它们就会越来越接近三角形，三角形的面积是很容易计算的。把这些“橘瓣”以相对的方向拼接在一起，所得到的形状就接近于宽为r、高为πr的矩形。即便你不是数学天才，也可以计算出圆的面积。

尽管这类方法最早是在古希腊时代提出来的，但是直到15世纪，德国哲学家尼古拉斯才用这个方法计算出我们所熟悉的πr^2。他认为，这个方法计算的其实是数量无穷多、面积无穷小的一个个图形的面积，因此不是一个严格的数学方法，不能得出精确的结果。但是，他承认这个方法可以有效地对正确答案进行预测，因为随着分割的图形越来越小，它们重新拼接而成的形状将越来越接近标准的矩形。

还有一些人也接受了这个方法，其中最著名的当属天文学家约翰尼斯·开普勒，但是，真正解决无穷小问题的却是与牛顿同时代的约翰·

沃利斯。如果不是被牛顿的光芒所掩盖，这位数学家的名气将会大得多。例如，他提出，我们可以认为那些用来计算总面积的小形状可以“稀释”，也就是说，它们可以根据需要变得非常小，但又不会彻底消失。这个词显然会让人们联想到牛顿取得重大突破时采用的那个方法。牛顿在提出流数术时，考虑的就是流动量（流数术这个名称由此而来）。牛顿的流数术不仅帮助他获知了万有引力的奥秘，还在数学家之间引发了一场持续百年的争论。

从艾萨克·牛顿的家庭图书馆的目录就可以看出，他的兴趣非常广泛。到他去世时，他拥有大约2 100本藏书。这在当时是非常了不起的藏书量，他所在的学院——剑桥大学三一学院所拥有的藏书还不到他的两倍。他有235本物理学和数学方面的书，有138本炼金术方面的书，神学方面的书也非常多，有477本。此外，他还有207部文学作品、46本游记、31本经济学著作，以及6本关于勋章的书（牛顿后来担任英国皇家造币厂厂长，负责铸造钱币和勋章）。由于兴趣广泛，相较科学研究，他花在炼金术与神学上的时间多得多。

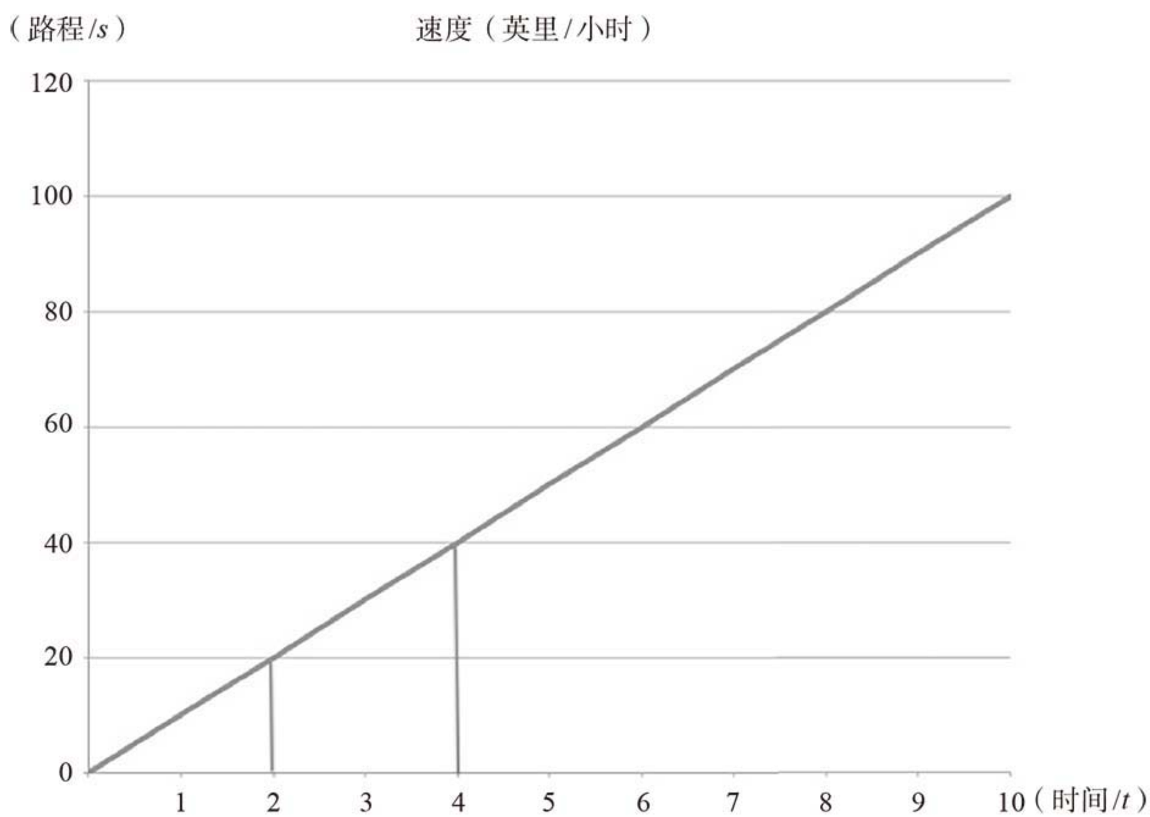
即便如此，牛顿仍然为科学做出了巨大的贡献，包括对光与颜色的研究和发现万有引力定律。他的力作《自然哲学的数学原理》是在流数术的基础上写成的，这个功能强大的方法从本质上看是一种数学工具，它给人一种违背自然规律的感觉，因为它的目的是预测宇宙间所有事物的行为。

牛顿晚年被奉为科学界第一名人，随之而来的是“牛顿神话”，人们认为牛顿在20岁出头的时候就发明了流数术。当时，由于爆发了一场瘟疫，剑桥大学疏散了校内人员，牛顿被迫回到林肯郡的家庭农场休假。在此期间，他进行了一番思考，但是，从他留下的笔记来看，牛顿的流数术显然是花费了20年的时间才逐渐成熟的。

牛顿的流数术也可以将图形分割成越来越窄的许多小形状，然后计算图形的面积。但是，牛顿提出这个新的数学方法的主要目的是，

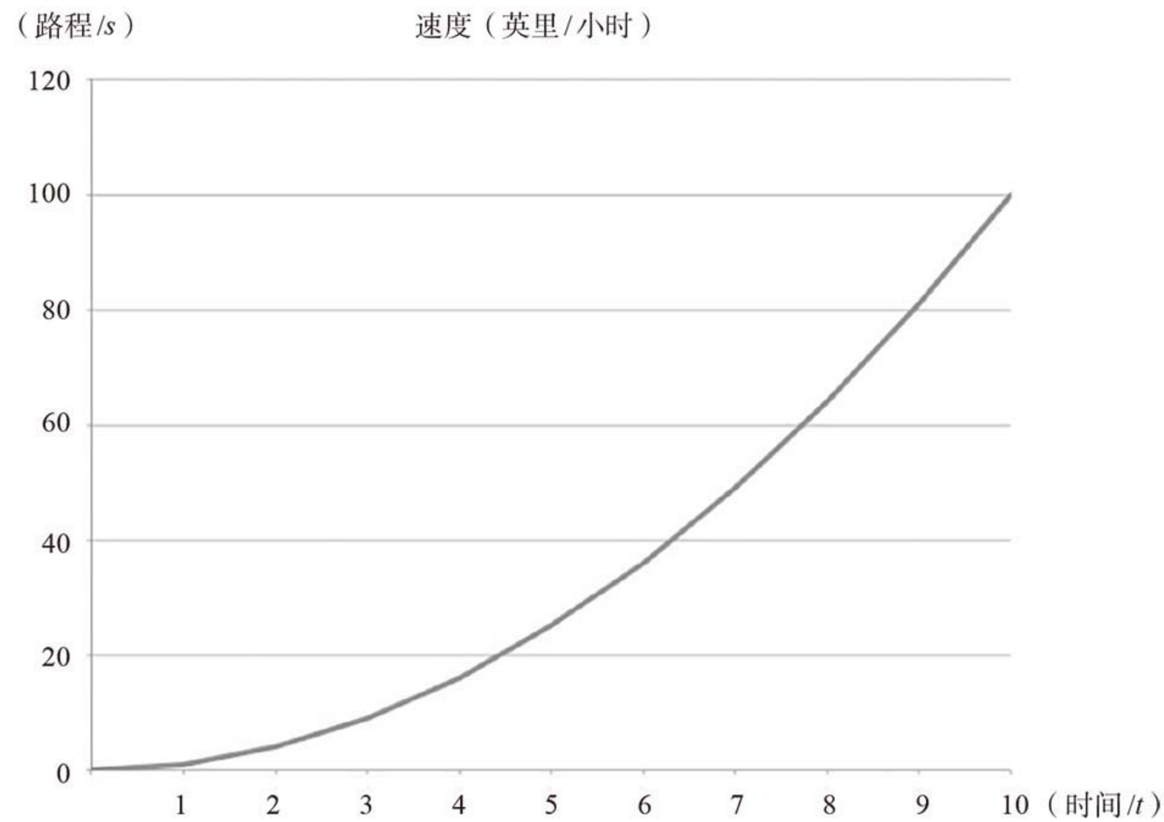
计算加速度等随时间变化的因素，这对于研究万有引力，以及比较月球绕地球运转与苹果自由落体运动之间的异同都具有非常重要的意义。利用牛顿的流数术计算加速度，就可以理解这个方法的原理。计算时需要使用几个方程，尽管这些方程非常简单，但是如果大家觉得有难度，可以跳过不读。

加速度是速度（包括速率和方向）变化量与时间的比值。为方便起见，我们在本例中假设运动方向保持不变，这样我们只需考虑速率的变化量。这是一种稳定的“线性”关系。假设1秒钟后的速度是每小时10英里，2秒钟后是每小时20英里，3秒钟后是每小时30英里，以此类推。为了计算出加速度，我们可以认为它就是一秒钟时间里速度的变化量。在本例中，速度每秒钟的变化量是10英里。研究这种加速度的一个便利方法，就是把它看作山的坡度。观察上图，就可以看出速度随时间的变化情况。



加速度就是这条直线的倾斜度，即斜率，是速度变化量与时间变化量之比。但是在现实世界中，很多关系并不像直线那样简单。例如，牛顿很早以前就知道，万有引力遵循平方反比定律，也就是说，万有引力的大小与物体和引力源距离的平方之间存在某种关系。把速度按照这种关系发生变化的情况绘制成图像，得到的将不是一条直线，而是一条曲线。

下面我们研究一下速度与时间存在简单平方关系时的加速度。在这种情况下，速度与时间的关系可以绘制成下面这个图。



由于该图不像直线那样便于处理，所以我们不能直接用速度变化量除以时间变化量来计算加速度。但如果我们考虑一个非常短的时间段，那么这段曲线就近似于一条直线。因此，我们可以利用速度变化量除以时间变化量的老办法，计算出一个加速度的近似结果。牛顿当年就是这样做的。我们用s表示速度，用t表示时间。牛顿把那个短暂

时刻里的微小时间增量称作“流动量”，并用一个类似于扭曲的零的符号（o）表示。因此，在这个短暂时刻里时间的变化量是（$t + o$） – t，时间变化又导致速度发生了变化（别忘了，$s = t^2$）：

$$(t + o)^2 - t^2$$

也就是说，我们可以用速度变化量除以时间变化量求出加速度：

$$\frac{(t+o)^2 - t^2}{(t+o) - t}$$

展开完全平方式，然后去掉括号，就会得到：

$$\frac{t^2 + to + ot + o^2 - t^2}{t + o - t}$$

上式化简后变成：

$$\frac{2to + o^2}{o}$$

分子、分母同时消去o，就会得到：

$$2t + o$$

最后，我们略去小小的流动量o，就会得到答案2。这个答案是正确的。至此，我们已经计算出当时间为t时，加速度是$2t$。但在得出这个正确答案的过程中，某些步骤有些不可靠，令人担心。既然o最终变成0，那么在前一步中，分子、分母同时消去o的做法就会涉及0除以0。我们已经知道，这种运算不符合数学法则。

牛顿非常清楚这个问题，因此他试图通过一种相当于约分的方式予以解决。他说，他处理的是比值，而且他的流动量就像约翰·沃利斯

说的那样，是可以稀释的，意思是流动量可以消失，但并不真的等于0。这个理由并不高明，但是流数术的确有效，牛顿也没有因为它不完全符合数学运算法则就放弃它。在《自然哲学的数学原理》一书里，他尽量将代数运算过程转换为几何证明，以便尽可能地掩饰这个问题，但是他很快就遇到一个更加迫切的问题：竞争。

竞争压力来自德国数学家戈特弗里德·威廉·莱布尼茨。据我们所知，在牛顿发明流数术的同时，莱布尼茨也在独立进行着同样的研究。牛顿知道德国人正在进行这项研究，因为他和莱布尼茨都与伦敦的英国皇家学会有联系。牛顿和莱布尼茨之间有过几次充满猜疑的通信交流，其中一封信非常有名，因为牛顿在信中写下了这句话：

> 我现在无法继续解释流数术，因此我宁愿将它隐藏在这段密码中：6accdæ13eff7i3l9n4o4qrr4s8t12vx

牛顿的做法是当时的一个学术惯例：用一句话概括流数术，然后把句子中字母的频率记录下来，用作保护自己成果的密码。牛顿这样做的目的是宣示自己的首发权，但根据解码后的那句话——*Data æquatione quotcunque fluentes quantitates involvente, fluxiones invenire: et vice versa*（已知包含若干流动量的方程，求流数；或者反过来，已知流数，求流动量），我们很难确定他的方法到底是什么。（后来，他又进一步做出了比较隐晦的澄清。）此时，牛顿完全可以公开发表他的方法并夺得首发权，但是直至去世，他也不愿意分享自己的想法，很多时候都是在同事的恭维下他才肯透露一二的。对于自己的成果，他总是秘而不宣。

1684年，莱布尼茨公开发表了他的研究成果，并为这项与流数术有异曲同工之妙的研究成果起了个名字——微积分。他的这个举动令牛顿对他的猜疑之心达到了顶点。莱布尼茨使用的是一套迥然不同的符号，而且实践证明这套符号使用起来更方便，但是它的原理与牛顿

的流数术如出一辙。牛顿使用的是一种点状符号，比如，在字母x上方加一个点，表示它的变化量，而莱布尼茨从希腊字母δ表示较小变化这个传统用法中得到灵感，用字母d表示被牛顿称为流动量的无穷小变化，即莱布尼茨的表示方法是dx/dt。这个符号不仅更清楚，而且处理起来更加方便。

莱布尼茨还为这种“微分”（按照尼古拉斯提出的办法，将许多小形状拼接在一起，以计算整个图形的面积）的逆运算设计了一个独特的符号。他把微分的逆运算称作积分，并用拉长的S——∫（来自拉丁语中表示求和的词）作为积分符号。牛顿的抱怨没有充分的正当理由，不发表研究成果也是他自己的决定。但是，由于牛顿的种种质疑，英国数学家都认为莱布尼茨有剽窃之嫌。

1708年，苏格兰数学家约翰·基尔在英国皇家学会的《哲学会刊》上发表文章，对莱布尼茨进行了直言不讳的指责（很有可能是牛顿要求他这样做的），气氛变得紧张起来。由于牛顿和莱布尼茨都是英国皇家学会的成员，因此英国皇家学会宣布成立一个11人委员会调查此事，以确定首发权的归属。由英国皇家学会会长亲自撰写的报告表明，情况对牛顿更有利。然而，这个结果并不令人感到奇怪，因为英国皇家学会当时的会长正是艾萨克·牛顿。从此以后，英国数学界与欧洲大陆数学界之间的关系降到了冰点，并且在随后几十年里都没有改善。

无论这两位数学家采取的是什么方法，无论到底是谁先创立了微积分，他们后来都遭到了哲学家乔治·伯克莱的猛烈攻击。这位主教在一篇文章中指出，牛顿和莱布尼茨在某些方面误入了歧途。这篇文章有一个气势恢宏的标题——“致分析家，或一位不信奉上帝的数学家”。伯克莱所说的那位不信奉上帝的数学家可能是指天文学家埃德蒙·哈雷，因为牛顿出版《自然哲学的数学原理》一书时得到了哈雷的帮

助。哈雷是无神论者，他对伯克莱的信仰提出了质疑。作为回应，伯克莱对流数术进行了讨伐。

伯克莱给流数起了一个富有诗意的名称——“逝去量的鬼魂”，并且指出，尽管流动量在计算过程中变成0，但后来它又回来了。要采用这样的方法，似乎必须借助信仰的力量，才能让人们接受一种无法想象的概念。伯克莱认为，那些批评宗教的人是表里不一的伪君子，因为他们的这个做法与宗教没有任何区别。无论这位主教是出于什么目的，我们已经知道，他提出的那个问题确实存在。的确，如果在运算过程中把那些无穷小量当作0来处理，流数术（也就是微积分）就是可行的，但这样的处理在数学上却是行不通的。

伯克莱称，牛顿和莱布尼茨能得出正确答案纯属运气好，是两个错误相互抵消的结果。他说：“接连出现两个错误，反而歪打正着得出了正确答案，但这不是科学。”牛顿的辩解理由是，这些极小的变化量具有流动性，它们并没有彻底消失，而是处于逐渐消失的状态。在上例中，当$2t + o$变成$2t$时，牛顿可以说结果趋近$2t$，是因为o趋近0，但是这个无穷小量永远不会真正等于0。从数学的角度看，这个解释只能算语焉不详、理由不充分的权宜之计。直到19世纪，才有两位数学家为微积分堵上了这个漏洞，彻底解决了这个问题。

19世纪20年代，奥古斯丁·路易·柯西重新定义了微积分中的无穷大和无穷小的概念，称它们是变量，意思是它们都趋近某个数值。随后，19世纪50年代，卡尔·魏尔斯特拉斯引入了极限的概念，这是我们今天仍使用的标准方法。有了极限的概念，我们就可以把某个终值确定为某个变量为无穷小时得到的极限值，条件是终值接近这个极限值的速度快于某个最低限度。魏尔斯特拉斯通过严格的证明告诉我们，只要接近极限值的速度足够快，微积分的方法就肯定有效。从某种意义上说，魏尔斯特拉斯在微积分计算的过程中抛弃了潜无穷的概念，而只要求接近极限值的（有穷）速度必须足够快。

我们将在第12章详细讨论无穷大这个概念，但是现在，我们只需知道这个概念在现实世界非常难以理解，甚至不可能被人们理解，然而，它对数学却产生了深远的影响。这个概念使牛顿的力学世界成为现实，从而证明了它具有无与伦比的价值。尽管极限概念的确立意味着无穷大永远失去了用武之地，但是微积分利用无穷小和无穷大搭建出的神奇的数学世界，就这样悄无声息地潜入了现实世界。在尘埃落定之后，人们惊奇地发现它与现实世界的关系竟然十分和谐。

微积分要求我们想象瞬间发生的情况，然而在极其短暂的瞬间，任何事似乎都不可能发生。古希腊哲学家芝诺提出的一个著名悖论——飞矢不动悖论，反映的就是这个问题。尽管下面的这个设想与飞矢不动悖论在文字表述上有所不同，但这是想象飞矢运动情况的最有效办法：我们可以想象一共有两支箭，一支飘浮在我们眼前的空中，静止不动，而另一支箭从弓弦上射出，闪电一般从第一支箭旁边飞过。

假设在第二支箭与第一支箭并排的一瞬间，我们冻结时间，然后研究这两支箭。在那一刻，这两支箭似乎一模一样。一支箭在运动，另一支则没有运动，但两支箭都悬浮在空中。芝诺说，我们不能区分两者状态的不同，说明我们对运动和变化的理解是非常片面的。

现在，我们知道这两支箭的物理属性在某些方面明显不同：那支运动的箭有惯性。尽管在时间静止的状态下我们无法感知这一点，但是惯性依然存在。此外，狭义相对论明确指出，物体在运动时，质量会变大。因此，如果古希腊人有能力，就可以比较两支箭的精确质量，从而区分它们的状态。

飞矢的这个比喻不是很好理解，但它还是说明了微积分的一个特点。尽管给人一种比较奇怪的感觉，但是微积分及其在探索自然时的应用方式都建立在现实世界的基础之上。如果我们可以处理时间上的瞬时概念和空间位置上的无穷小变化的概念，微积分就是一种非常自

然的数学方法。毕竟，它不是晦涩难懂的数学研究的产物，而是在不断拉近视野，研究世界上事物的运动情况的过程中获得的成果。

从某种意义上说，微积分的研究对象正是牛顿的“机械宇宙观”。受到它的启发，拉普拉斯设想，如果我们掌握了足够多的信息，就可以根据过去预测未来。但在牛顿完成他的研究之前，已经有人埋下了种子，希望可以借助数学对未来进行不太确定的预测——基于机会和条件的预测。

---

1. 皮埃尔–西蒙·拉普拉斯（1749—1827），法国天文学家、数学家，法国科学院院士，天体力学的主要奠基人。——译者注

第10章
卡尔达诺：概率与“水晶球”

Are Numbers Real?

统计学作为一门学科，它的历史可追溯至史前人们掰手指数山羊时。如果我把记录邻居每次借山羊数量的符木都保留下来，稍加比较，就可以发现邻居借山羊的习惯随着时间变化而有所波动。也许我能做的只是进行简单直接的比较，看看数量的变化，但是这种基本的统计工作仍然会让我乐此不疲。

统计学（statistics）这个词与“国家”（state）源于同一个表达，这门学科在刚开始时只是指收集一个国家的相关数据，这与美国中央情报局的《世界各国概况》没什么两样。显然，这样的活动不会引起任何麻烦。但是，只要统计学继续存在，有一句名言就会如跗骨之疽，让它笼罩在阴影之中：“世界上有三种谎言，分别是谎言、讨厌至极的谎言和统计数字。”这显然是指责统计人员居心叵测，尽管他们可能以无辜的数学家自居。

我们都不清楚，这句将谎言与统计数字相提并论的话到底是谁说的。通常认为，这句话出自英国前首相本杰明·迪斯累里之口。但是，这位擅长以诙谐幽默的语言讽刺他人的前首相却矢口否认，声称他说

的这句话引自马克·吐温，但是人们在马克·吐温的作品中却找不到相关证据。也许是这位小说家造访英国时，随口跟这位政治家说的吧。

然而，在刚开始的时候，统计学这门学科的确给人一种恐怖的感觉。第一位统计人员的统计对象是死亡人数，这位统计人员名叫约翰·格朗特，是一名纽扣制造商。尽管他从事的工作与数学无关，但是他对周围世界的运行规律颇感兴趣。格朗特想办法收集“死亡公报”以了解伦敦1604—1661年死亡人口的详细资料，还收集了出生人口数据，然后将这些数据汇集成册。他的目的是通过研究这些数字，了解伦敦底层人民的生活概况。

在一定程度上看，他的工作就是收集散落于各种文件中的相关数据，然后将这些已有数据公布出来，这是人类有史以来第一次了解到不同年份瘟疫致死人数的情况。然而，仅仅整理这些已有的数据，并不能让格朗特感到满足。他还将数字加以整合，从而发现前人没有发现的信息。例如，他根据整合后的数据估算出伦敦的人口数量（当时还没有人口普查），并试图了解不同人群的预期寿命差异。

正是这项预期寿命研究，再加上天文学家埃德蒙·哈雷后来所做的分析，直接催生了一个新的行业——保险业，这个行业主要针对的是人们在综合考虑统计数据与未来可能情况之后的不确定心理。当时，人们喜欢聚集在伦敦的咖啡屋里谈生意。这种把赌注押在未来结果上的行业就始于这些熙熙攘攘的人群，随后，借助当时最全面的统计数据，迅速向世界各地蔓延，成了每个人都要打交道的一个行业。

尽管遭到了迪斯累里的鄙视，但是统计学作为一门独立的学科，发展的态势似乎一帆风顺。当统计学与概率（研究可能性的数学分支）相遇之后，更是迸发出耀眼的火花。在此之前，数学在它与现实世界之间的关系中一直处于从属地位，它诚实地展示当前的状况或者解释已经发生的事。但是，统计学这个全新的数学分支却在社会底层人民的支持下，大言不惭地预测起未来，而且它的预测结果与牛顿

的“机械宇宙观”不同，充满了不确定性和风险。于是，数学描述周围世界的能力取得了重大突破，并把触角伸至尚未发生的未来。最终，概率和统计学的重要性与日俱增，被用来描述包括气体特性、神秘量子在内的所有事物。我们将在第13章深入讨论这方面的内容。

要想在蓬勃发展的保险业叱咤风云，仅掌握足够多的数据是不够的，还必须把这些数据变成“水晶球”（格朗特已经证明，这是有可能的），才能用它们预测未来。所谓水晶球，是指那些社会底层人民的生活习惯，而不只是那些纽扣制造商的各种癖好。这是一个赌徒的世界。仔细想想，保险业就像一个赌场：它披着行业的外衣，希望可以通过赌局保持不断前进的态势，“玩家”虽然有赚钱的可能性，但在大多数情况下，他们投入的钱都变成了保险公司的利润。

赌博业的历史源远流长，人们曾经在有几千年历史的考古地点发掘出几个表面光滑的指关节，这是一种早期的四面体色子。自从有了硬币之后，抛硬币的游戏就开始兴起，而且似乎经久不衰。这个游戏非常简单，可以直接利用硬币的正反两面得到随机数据。至少在硬币没被动过手脚的情况下，它产生的都是随机数据。自古以来，人类就嗜赌，无论是对赛跑还是天气打赌，都能让人们享受到赌博的乐趣。总体而言，无论是赌徒还是诚实的庄家，他们依赖的基本都是直觉和猜测。然而，在意大利数学家（也是一名狂热的赌徒）吉罗拉莫·卡尔达诺出现之后，这种状况就发生了彻底的变化。

前文在讨论虚数时提到过卡尔达诺，他把变幻莫测的可能性引入数学世界，这不仅对数学的未来发展具有重要意义，而且在将数学与难以捉摸的寻常事物分离开的过程中，也发挥着举足轻重的作用。卡尔达诺出生于1500年前后，20多岁时开始撰写一本关于概率的书，但是直到他60多岁时才写完。最终，这本书于17世纪60年代出版。一般而言，这么长的时间跨度足以让它淡出人们的视线，但是它的出版仍

然引起了人们的广泛关注，这说明卡尔达诺的思想非常超前。这本书就是《机遇博弈》（*Liber de Ludo Aleae*）。

几乎所有经常玩抛硬币游戏的人都知道，如果硬币没有问题，抛出正面和反面的概率是相等的。没有人知道下一次抛掷会出现什么结果，但是出现正面或者反面的可能性是均等的。卡尔达诺的贡献在于，通过简单直接的观察，将观察结果变成一个数字结构——把分数的概念与对未来的预测结合起来，使我们对一个简单的系统（例如抛硬币）有了深刻的理解。

当然，硬币没被动过手脚这个限制条件非常重要。要让人们尊重概率，难点之一就在于赌徒（尤其是职业赌徒）经常作弊。有的职业赌徒通过使用两面都是正面图案的硬币，在抛硬币游戏中无往不利，有的则在三牌赌皇后游戏中熟练使用简单而有高度欺骗性的"从最上面拿牌"的手法[注]。但是，他们都有一个共同点：他们仿佛具有某种魔力，可以轻易地误导那些容易上当的对象，说服他们参加赌博游戏。我们至少可以认为职业赌徒、魔术师和小偷之间的界限比较模糊。

我在做关于概率和统计学的报告时，通常会一开始先举抛硬币的例子。我会拿出一枚硬币，并告诉观众，我在报告开始之前已经花了一些时间抛这枚硬币，并且最后9次的结果都是正面。（这个结果完全是有可能的，但是通常需要花一点儿时间。）然后，我问观众，如果我再抛一次硬币，会出现什么样的结果？一种观点认为，既然正面和反面各有一半的概率，在出现这么多的正面之后，下一次出现反面的可能性应该更大。还有一种观点认为，因为这枚硬币明显偏向正面朝上，因此下一次出现正面的可能性更大。到底哪一种观点是正确的呢？总有一些人会说，"出现反面的可能性更大"，这就是所谓的"赌徒谬误"，因为在现实世界中，硬币没有记忆能力，之前的结果不会对之后的结果产生任何影响。然而，在连续出现同一个结果之后，人们很容易就会以为接下来出现另外一种结果的可能性更大。

通常情况下，大多数观众都会给出正确答案：出现正面和反面的概率各一半。但是，有一些人仍然认为出现正面的概率更大。这可能是经常出现在体育比赛中的另外一个谬误——热手谬误。所谓热手谬误是指，体育迷认为一连串好的结果意味着某位选手或某支球队将保持“连胜势头”。但如果我又连续抛出三个正面，观众就开始产生怀疑。他们的怀疑是正确的：我使用的硬币两面都是正面。（此时，观众会提出相同的问题：“你从哪里搞到这枚硬币的？”答案是电子港湾网站。）有趣的是，观众不可避免地对这枚硬币产生了强烈的兴趣，就好像电影中的高明骗局使我们欲罢不能一样。他们希望看一看这枚有两个正面的硬币，还想亲手摸一摸这个邪恶的道具。

在卡尔达诺那个年代，人们都知道，只要硬币质地均匀，出现正面或反面的可能性是一样的。（严格地说，真实情况并非如此。根据抛掷的方式，标准硬币出现正面和反面的概率大约是51∶49或49∶51，第一次抛掷时朝上的一面略占优势。）但是，没有人把这种机会均等的情况变成一种适合数学研究的形式。尽管表达抛硬币时正面朝上的可能性的方法有很多，诸如机会均等，各占一半，但是只有用可以进行算术运算的数字来表示，它才最有利于数学研究。第一个提出用从0（表示“不会发生”）到1（表示“肯定会发生”）的数字表示概率的人正是卡尔达诺。根据这种方法，硬币正面朝上的可能性可以表示为1/2。

这种表现形式直截了当，但是除了为预测行为奠定数学基础之外，卡尔达诺还有其他的贡献。［卡尔达诺应该没有使用“概率”（probability）这个词。从14世纪开始，法语中就出现了这个词，意思是“不确定，但是有可能”。至于具有现代数学意义的“概率”概念，最早的使用记录只能追溯至1692年。］套用他处理抛硬币时使用的那个方法，我们可以说，从我们现在使用的一副普通扑克牌（不包括大小王）中抽出某一张牌的可能性是1/52。

卡尔达诺还提出了计算组合概率的两个重要方法。后来的事实证明，这两个方法对于所有赌博游戏玩家来说都具有非常重要的意义（别忘了，卡尔达诺不仅是一名数学家，还是一名狂热的赌徒）。第一个方法可以帮助我们计算得到多个可能结果的组合概率。比如，根据卡尔达诺最初的理解，我们知道掷一次色子得到任何特定点数（例如6）的概率是1/6。但是，如果你想知道得到1点或者6点的可能性，答案就应该是2/6，也就是1/3。

卡尔达诺还证明，计算两枚色子掷出相同点数（例如在双色子游戏中掷出两个6点或者两个1点）的组合概率的方法是将两个分数相乘，也就是1/6×1/6，即1/36。因此，得到某个相同点数的可能性只有1/36。此外，他还发现，这与用两枚色子掷出一个1点和一个6点略有不同。要得到后面的结果，一共有两种方法：第一枚色子得到1点，第二枚色子得到6点，或者第一枚色子得到6点，第二枚色子得到1点。因此，概率是1/36 + 1/36 = 2/36，即1/18。

卡尔达诺最巧妙的一个发现是计算双色子游戏中任意一枚色子得到6点的概率。也就是说，我掷两枚色子，至少有一枚掷出6点。至于是一个6点还是两个6点，以及哪枚色子得到6点，我都不在乎。我们经常会遇到这种组合概率，而人们的自然反应是使用加法。每枚色子得到6点的概率都是1/6，因此第一反应是把它们加到一起。但这种做法显然是错误的，否则，只需6枚色子就能确保得到一个6点。而玩过色子游戏的人都知道，真实情况并不是这样。

现在的问题是要想办法表示“任意一枚”的可能结果。卡尔达诺的高明之处在于他发现，这个问题可以先转化为“两枚色子都没有”的问题，再用他发明的方法，即用乘法算出概率。如果一枚色子得到6点的概率是1/6，那么结果不是6点的概率就是5/6。因此，两枚色子都没有掷出6点的概率是5/6×5/6，即25/36。也就是说，两枚色子中有任意一枚掷出6点的可能性是1 – 25/36，即11/36。与用一枚色子掷出6点的概

率相比，前者比后者的两倍（12/36）还小。随着色子的数量增加，这个概率将会趋近1（也就是肯定有色子掷出6点），但永远不会等于1。因此，即使同时掷出很多枚色子，也有可能没有一个6点。

在卡尔达诺之后，人们对他的研究成果进行了完善和发展，其中最著名的是法国数学家布莱瑟·帕斯卡和皮埃尔·德·费马，两人合作解决了一个众所周知的难题，从而让概率变成一个深受保险业欢迎的工具。他们解决的那个难题叫作"点数分配问题"。两名势均力敌的玩家因为一笔奖金而"激战"，根据规则，点数首先达到某个数字的玩家获胜。但是，如果他们在游戏结束时还没有决出胜负，该怎么分配那笔奖金呢?

假设每赢一局就得一点，在游戏结束时，一位玩家有12点，另一位玩家有7点。帕斯卡认为，要想合理地分配这笔钱，就需要考虑若游戏可以一直持续下去直至两人决出胜负，每名玩家需要赢多少局才能获胜。假设设定的目标是15点。在这种情况下，第一位玩家只需再赢3局就可以获胜，而第二位玩家还需要再赢8局。帕斯卡根据双方获胜还需要赢得的点数，考察了接下来可能发生的情况，然后用数学语言给出了一个公平分配奖金的方案。他提出的其实是一个叫作"期望值"的概念。所谓期望值，是指根据预期，某个可以产生随机结果的过程在连续重复多次后可能得到的结果。

下面我举一个非常简单的例子。假设游戏规则要求你连续掷色子10次，然后根据掷出的平均点数获得相应的现金。赌注设为多少时，这个游戏才值得参与呢？常识告诉我们，我们赢到的钱可能是概率的中值。难得的是，这次我们的常识是正确的（在涉及概率时，常识往往并不可靠）。你也许会不假思索地回答3，因为3是6的一半。但是，如果我们把1—6这6个值排成一排，就会发现中间值应该是3和4的平均值，也就是说期望值是3.5。

我们也可以通过一种更严谨的方式来考虑这个问题。掷出1点的可能性是1/6，掷出2点的可能性是1/6，以此类推，掷出6点的可能性也是1/6。求1×1/6 + 2×1/6 + 3×1/6 + … + 6×1/6的和，得数为21/6，即3.5。既然你有可能赢得的预期奖金是3.5美元，那么赌注低于这个金额都是可以接受的。在任意一局中，你都有可能输钱，但是只要玩的局数足够多（并准备足够多的本金），最终的赢家应该还是你。

计算交易期望值的概念绝不仅限于赌博，它是各种现代金融系统的基础，其中最典型的例子就是保险公司。它们就像赌博玩家。保险公司通过设定赔率，保证即使自己在某一“局”（他们称之为“保险单”）赔钱，也总体来说一定会赚钱。当然，赌场也是这样。重要的是，这个计算方法可以用于权衡不同的选择方案，并帮助我们做出最有利的决定。

比如，假设你有两个可能的投资方案。一个投资方案有1/2的可能性赢利1 000美元，有1/2的可能性不赢利；另一个投资方案有1/4的可能性赢利1 900美元，有3/4的可能性不赢利。哪个投资方案更有利呢？我们可以用概率乘以投资结果的方式计算出期望值。如果选择第一个投资方案，期望值就是500美元，而第二个方案的期望值是475美元。因此，第一个投资方案对你更有利，尽管第二个方案有可能赢利更多。如果某个投资方案会产生不止一个可能的结果，就要把发生这些结果的可能性加到一起。

同其他基于概率的预测方法一样，期望值也没有魔力，无法完成不可能的任务。期望值不会告诉你掷一次色子能赢得什么，但是只要你掷色子的次数足够多，就可以根据期望值预测可能的结果，至少在公平游戏中可以做到这一点。伯努利家族的一位才华横溢的成员指出，在某些情况下，期望值也不可靠。

在介绍伯努利的发现之前，我们先设想一种十分荒谬的彩票，以此说明期望值这种简单的计算方式有时未必有效。（我之前举的例子

都是碰运气的游戏，在这些游戏中我们可以计算出精确的概率。同样的方法也可以应用在商业投资、购买保险等方面，但是此时，我们只能根据具体情况对概率做出估计。）

这种彩票有两种票面，价格都是10美元，但是第一种票面有9/10的概率赢得11.11美元，而第二种票面有1/100 000的概率赢得100万美元。所以，这两种票面的期望值都是10美元。期望值与票面价格相同，对于彩票而言是非常难得的。在彩票与赌场等赌博游戏中，期望值通常必须低于票面价格，这样经营者才有利可图。但是，这种彩票的经营者非常慷慨。因为这两种票面的期望值相同，所以我们在购买彩票时应该不会过于关注选择哪一种。但是，这两种票面带来的结果似乎大不相同。结果是否诱人，决定因素似乎不是期望值，而是你的个人情况。到底选择哪一种票面，可能要看10美元在你的日常生活中具有什么样的意义。

为帮助大家更好地理解这一点，我举一个更夸张的例子。我在讲座中谈到我的《色子世界》这本书时，经常会跟观众做一个叫作“最后通牒博弈”的心理游戏。心理学家经常通过这个实验告诉大家，经济学家根本不了解人的心理（心理学家都喜欢揭经济学家的短儿）。通常，这个游戏会设立一笔小奖金（例如1美元），由两名玩家展开博弈。第一名玩家告诉第二名玩家这笔钱的分配方案，第二名玩家可以说“行”或者“不行”。如果第二名玩家说“行”，这笔钱就会按照第一名玩家制订的分配方案进行分配。如果第二名玩家说“不行”，那么他们两个人都不会有任何收获。

经济学家和逻辑学家都认为，只要第一名玩家不打算独吞这笔钱，第二名玩家就会接受他提出的任何分配方案，因为拒绝接受意味着一分钱也拿不到，这样的决定似乎太不合理了。你可以问任何人一个问题：“如果有人白送你一些钱，你会拒绝吗？”答案通常是：“当然不会！”但是事实上，如果第一名玩家分给第二名玩家的钱低于奖金总

额的30%，第二名玩家通常就会拒绝接受。这个数字适用于美国人和欧洲人。不同国家的人对分配方案有不同的要求，但是绝大多数人都对分配比例有一个最低要求。为了惩罚另一位玩家的不公平做法，人们宁愿承受一定的经济损失。但我们也可以利用这个游戏，反过来证明心理学家对人们心理的把握也不是很准确。

在玩完传统意义的最后通牒博弈游戏后，我请参加讲座的观众在脑海里重玩这个游戏，但这次的奖金不是心理学家提供的，而是一位大富豪，奖金额增加至1 000万美元。（事实上，我在做这个实验时，通常会把奖金设为1 000万英镑，但结果没有什么不同。）现实点儿说，如果第一名玩家分给第二名玩家10万美元，第二名玩家很可能不会拒绝，尽管他只能得到总奖金的1/100，而第一名玩家能得到990万美元。因此，我让观众都站起来，然后按照由多至少的顺序，告诉他们可以从这1 000万美元中分得的金额。我还告诫他们要诚实，一旦觉得我给出的金额低于他们愿意接受的最低值，就坐下来。

做实验时，我们使用的不是真钱，因为我仍在苦苦寻找愿意资助这项实验的大富豪。我觉得，由于不是真金白银，很多人夸大了他们拒绝接受的金额。但是，通常情况下，在金额高于50 000美元时决定坐下来的人不是太多；在金额降到10 000美元以下、5 000美元以上时，大多数观众都会坐下来；等到金额降至500美元时，站着的人已经寥寥无几了。当我说出1美元时，只有1—4名观众仍然站着。一想到人们为了报复对方而宁愿放弃（至少他们声称如此）一大笔钱，我就觉得这个实验非常有意思。我在前面介绍的那种奇怪的彩票，给了人们两个选择：一个是有9/10的概率赢得11.11美元，另一个是有1/100 000的概率赢得100万美元。结果，人们的反应与他们在最后通牒博弈游戏中的表现是一样的。在最后通牒博弈游戏中，最后仍然站着的人通常是青少年。1美元对于他们的意义远胜于在中年观众心目中的价值。

说到这里，我们回过头去介绍伯努利家族的那名成员，看看他对期望值概念的缺陷有哪些认识。这名成员就是数学家尼古拉斯·伯努利，他是约翰·伯努利的儿子，丹尼尔·伯努利的弟弟。（在这个成就显赫的瑞士家族中，丹尼尔的名气最大。）尼古拉斯研究过一个简单游戏的结果，在这个游戏中，我们需要做的就是记录抛硬币得到的一系列结果。玩家能赢多少钱，取决于他抛硬币的结果。只要抛出反面，奖金就会加倍，游戏继续进行。一旦得到正面，游戏立刻结束，玩家的收获只是当时的奖金。

假设我们开始时的奖金是1美元。如果第一轮抛硬币的结果是正面，你就会赢得1美元；如果是反面，奖金就会加倍，而且你可以再抛一次。如果第二轮的结果是正面，你就会赢得2美元。如果你坚持到第三轮且得到的结果是正面，你就会赢得4美元。如果你第三轮得到反面，并且第四轮的结果是正面，你就可以赢得8美元，以此类推。尼古拉斯指出，最有意思的是，把奖金定为多少，你才愿意参加游戏？我们应该采取的做法是计算期望值，如果奖金低于期望值，就值得参与。

要计算出期望值，我们需要知道每次抛硬币时第一次出现正面的概率，然后用它去乘以此时的盈利，再把所有可能的结果加到一起。第一轮抛硬币时，得到正面的概率是1/2。在这种情况下，奖金是1美元，它贡献的期望值是1/2×1美元= 0.5美元。第一轮得到反面且第二轮得到正面的概率是1/2×1/2 = 1/4，此时的奖金是2美元。因此，它贡献的期望值是1/4×2美元= 0.5美元。第三轮得到正面的概率是1/8，奖金为4美元，期望值是1/8×4美元= 0.5美元。我们已经可以看出其中隐藏的规律了：每一轮的期望值都是0.5美元。

因此，只要把所有可能盈利的期望值加在一起，就可以计算出总期望值。也就是说，总期望值为：

（1/2×1美元）+（1/4×2美元）+（1/8×4美元）+（1/16×8美元）+…

=0.5美元 + 0.5美元 + 0.5美元 + 0.5美元 + …

别忘了，“…”表示继续下去。因此，上面的计算结果表明，无论参加这个游戏需要投入多少钱，根据期望值，你都应该参加。例如，即使参加这个游戏需要投入100万美元，你也应该参加，因为0.5美元 + 0.5美元 + 0.5美元 + 0.5美元 + …的值大于100万美元，实际上，这个和比任何数都大。这个级数的极限是无穷大，也就是说，这个游戏的期望值是无穷大。但是，尼古拉斯·伯努利强调的问题是，只在同样的过程重复很多次时，期望值才真的有效。对于具体某一轮的情况，期望值的效果就不那么好了。

很难想象有人愿意拿出100万美元，去玩一个只有50%的概率赢得1美元的游戏。其实，玩家只要想一想他们输钱的可能性，就会知道该怎么做。例如，我们知道，最多赢1美元的概率是50%（即1/2），最多赢2美元的概率是75%，最多赢4美元的概率是87.5%，最多赢8美元的概率是93.75%，最多赢16美元的概率是96.875%。也就是说，即使拿出16美元参加这个游戏，亏本的风险也很大。

因为觉得好玩，我刚刚又玩了一次抛硬币的游戏。想知道结果吗？我在第三轮抛出了正面，这意味着我可以赢得4美元。所有利用概率工具的人都要注意使用它的场合。计算两枚色子中至少有一枚色子掷出某个点数的概率并没有多大的难度，但是我们经常需要这种能力（例如在玩双陆棋时）。当我们在权衡是否要做出某种经济上的承诺时，我们也经常需要计算期望值。但是，考虑是否在某个游戏或投资活动中使用概率工具，不能仅依据“大量人口的平均结果”或“大量交易的平均情况”，还要考虑它可能造成的后果。

例如，某个银行系统通常运行顺畅，但是若每完成10 000次交易就把账户数据全部清除，我们对此肯定无法接受。如果你的账户正好是那个不幸被清空的账户，即便这套系统完美地完成了99.99%的交易，也无法平息你的怒火。因此，即使99%的案例都得到了妥善处

理，性能统计的结果仍然取决于那些处理不当的案例会造成什么样的后果。如果是快餐店没有按时送来汉堡这种无关痛痒的事，这样的统计结果不会导致大问题。但如果是医院出具的常规体检报告中的死亡可能性，就肯定会让人十分担忧。

实践证明，在涉及大量数据或者大量调查对象时，基于概率的统计可以发挥极其重要的作用。无论这些调查对象代表的是“美国人民”还是“汽缸中的气体分子”，只要可以忽略统计方法对个体造成的影响，我们就可以借助数学工具对这些调查对象代表的群体行为做出准确程度较高的预测。

苏格兰物理学家詹姆斯·克拉克·麦克斯韦是在科学研究中最早大量使用数学工具的物理学家之一（我们将在下一章深入讨论），也是最早运用统计方法研究气体属性的物理学家之一。他选择了一些有强烈气味的物体作为研究对象。这些物体的气味（难闻或者好闻并不重要）传播到人的鼻子里，为什么需要那么长的时间呢？在19世纪，人们普遍认为气体分子的传播速度非常快，每秒可以运动几百米（或几百码）的距离，但是气味通常需要几秒钟的时间才能充斥整个房间。

德国物理学家鲁道夫·克劳修斯认为，这是因为分子发生碰撞的缘故。虽然分子的运动速度的确非常快，但是它们彼此之间不停地碰撞，以致改变了运动方向。所以，一堆新的分子（“气味分子”）需要很长时间才能完全扩散到空气中。

克劳修斯认为所有气体分子的运动速度都一样。但是，麦克斯韦认为这个说法没有道理，他更倾向于气体分子的运动速度各异，有的较快，有的较慢，速度分布曲线的峰值在某个区间范围内。麦克斯韦认为，如果确实如此，那么只有借助统计法，才能全面了解气体分子的特性。这就是所谓的“麦克斯韦分布”。尽管气体分子的运动速度随温度的变化而变化，但是麦克斯韦分布却找到了一个可行的计算方法。从此以后，人们掌握了预测气体变化特点的能力。

这种通过统计掌握多个变化个体的普遍情况的能力，不仅可以用来研究分子的特点，还可以用来研究人的行为。掌握了这种技能之后，我们才有可能了解大型人群内部正在发生的变化，并完成各种各样的预测，例如服装销量、药品需求等。但我们必须清楚，它也有局限性。即使是分子的统计特性，也有可能造成误导性的结果。我们以热力学第二定律为例。该定律称，热由高温物体向低温物体传递，封闭系统中的无序状态会保持不变或者增加。人们往往认为这是一条颠扑不破的真理，但事实上，它也是建立在统计学的基础之上的。

比如，根据这条定律，如果我们将两个盒子之间的隔板去掉，经过一段时间之后，两个盒子中温度不同的气体将混合到一起，变成均匀气体，其温度介于之前的两个温度之间。这是根据热力学第二定律得到的结果（两组有序程度较高的分子通过温度的选择，变成了无序的混合体）。但是，从理论上看，这些气体有可能是在重新建立短暂的完全随机的温度阶梯。一个盒子中的高温分子有可能碰巧比另一个盒子多，由于分子的数量非常多，这种偶然性不大可能产生非常大的影响，但是这种情况的确有可能发生。统计数据表现的是总体可能性，而不是必然性。

在使用统计方法研究人的活动时，我们有可能把典型规律套用到独特群体（例如上文所说的那些高温气体分子）上，还有可能认为关于一群人的统计规律适用于某一个体。我们无须考虑气体中单个分子的特性，因为所有分子基本上都是相同的，但人与气体分子不同。统计学历史上有一个非常有名的案例。1999年，一个英国母亲萨莉·克拉克被判定杀死了她的两个幼子，并因此在监狱中服刑近4年时间，直到这项判决被推翻之后才重获自由。克拉克含冤入狱的原因是，法庭在运用统计学工具时犯了严重的错误，不但相关人员的计算能力不过关，他们还将统计得出的整体普遍情况与个体的特定情况混为一谈。

这次审判是在克拉克的第二个幼子死亡之后进行的。造成克拉克的两个儿子在不足3个月时就夭折的罪魁祸首是婴儿猝死综合征（SIDS）。著名儿科专家、教授罗伊·梅多爵士应检方邀请，作为专家证人参与了此案的审判。不幸的是，梅多在概率与统计学方面的知识并不全面。研究表明，在没有其他影响因素的情况下，一个家庭中发生婴儿猝死的概率是1/8 543。梅多告诉陪审团，克拉克的两个儿子都死于婴儿猝死综合征的概率是这个数的平方，约为1/73 000 000。梅多声称，这种情况堪称百年不遇。

这个证据在克拉克案的判决中发挥了重要作用，但是其中存在着巨大的错误。卡尔达诺早就发现，两个无关事件的组合概率的正确计算方法是乘法。因此，我们知道，用一枚色子掷出6点的概率是1/6，连续掷出两个6点的概率是1/6×1/6 = 1/36。两次投掷是彼此不相关的两个事件，即第一次投掷不会对第二次投掷的结果产生任何影响。

但是，这次审判却忽略了一个问题：这个数学工具并不适用于婴儿猝死的情况。有充分的证据表明，这两起婴儿死亡事件并非彼此无关。如果一个家庭中发生过婴儿猝死，那么这类事件再次发生的可能性要远远高于普通家庭发生婴儿猝死的可能性。真相澄清后不久，有人公开发表研究结果，称一个英国家庭发生两个婴儿猝死事件的可能性并不是百年不遇，而是每18个月就有可能发生一次。

除了概率计算的错误，这起案件还存在其他问题，他们误认为“某件事偶然发生的概率很低”的意思等同于“发生某件事的原因并非偶然”。这是一个巨大的逻辑错误，没有任何道理可言。以“欧洲百万”乐透游戏为例，虽然中大奖的概率仅为1/116 531 799（这个可能性低到令人发指的地步），但是在大多数情况下，每周都会有幸运儿产生。由于参与的人非常多，所以不可能发生的事件也会时不时发生。我们并不能因为某件事发生的可能性非常低，就认为它不会发生。

即便涉及的人群不大，低概率事件也可能会发生。除忽视了这个简单的事实以外，他们还犯了一个严重的错误。婴儿猝死的概率是1/73 000 000（这个概率其实并不准确），但是他们居然因此认为，如果73 000 000个婴儿中的其他72 999 999个婴儿死亡，就一定是死于谋杀。检方应该认真权衡的两个概率是婴儿猝死的概率与一个英国家庭中母亲连续杀死两个孩子的概率，后者肯定不是72 999 999/73 000 000。统计学可以在科研（和法律事务）中发挥巨大作用，但前提条件是要正确使用。

显然，适用于热力学第二定律的统计学工具不能用来研究人的行为，尽管这个想法颇具诱惑力。人们倾向于“集体审议”或者聚众闹事这类行为，尽管其中有心理学原因，但这也说明一群人聚在一起，其行为的复杂程度不是一堆气体分子可以比拟的。在美国科幻小说家艾萨克·阿西莫夫以心理历史学的概念为基础创作而成的“基地”系列小说中，人们可以用一种异常强大的统计学工具研究某个文明，预测它未来的发展情况，甚至可以具体到某个事件。但是，现实中却永远不可能找到这样的统计工具。

阿西莫夫的创作灵感来自爱德华·吉本的经典历史学著作《罗马帝国衰亡史》。这部著作似乎告诉读者，人们可以从具体事物中找到线索，从而预见一个帝国走向衰亡的结局。阿西莫夫把这个（本来就不可靠的）概念拓展成一个研究行为特性的纯粹的数学工具。但是，为人们在现实中利用统计数据来预测复杂事物（例如文明）的未来情况时，就会像那些试图进行长期天气预报的人一样，面临同样的困难：系统过于复杂，涉及的变量过多，以致难以做出有意义的预测。在数学上，这被视为一种高度混乱的情况。这意味着开始时发生的小变化（在研究对象是一群人时，这些小变化通常是由个人的行为引起的）会对结果造成巨大的影响。

概率和统计学已经成为许多科学家手中威力巨大的武器。但是，事实证明，如果这些科学家的数学造诣不深，滥用统计工具就会造成一系列问题。毫无疑问，数学很有用，在科学研究中可以发挥重要作用。但是，如果过于重视统计学的“证据”作用，不仅对科学研究没有任何益处，还会导致我们在得到看似正确的数据之后做出错误的判断，还自以为揭开了天地万物的奥秘。

有的问题根本不是数学的错，而是数学工具应用不当造成的。我在《超感官》一书中讨论的超自然研究就经常犯这样的错误。假设我们正在测试心灵感应能力。我们预先安排了一场选拔测试，得分高的人才能留下来，成为实验对象。只要我们在正式测试时不考虑他们选拔测试的分数，这种做法就无可厚非，但是，选拔测试的得分通常会被计入正式测试的成绩。这些人之所以能入选，是因为他们的选拔测试得分很高，因此他们肯定会使测试结果偏向肯定性的一面。

这种“摘樱桃”式的有选择性地使用数据的行为，是统计中的一个常见错误。如果只选择那些对假设有利的数据，忽略或者不重视其他数据，最后得到的结果就会毫无用处，但这种情况却经常发生。有时候，这是一种明目张胆的有意行为；有时候，例如选拔测试，则是无意行为，测试者甚至不知道他们的行为已经导致结果发生了偏差。还有一种可能的情况是想方设法舍弃一些数据。如果实验出了问题，那么在审核这些数据之前将它们舍弃就不会有任何不妥。但是，数据一经审核就不应被舍弃，否则就会有选择性使用数据之嫌。有时候，我们甚至会下意识地找一个理由，去舍弃那些不利于预期结果的数据。

早期心灵心理学实验者J. B. 莱因也犯过一个无意识的“摘樱桃”的错误。莱因在历时多年的实验中，利用一套“齐纳”牌，针对多名个体进行了多次心灵感应测试。这套牌共计25张，每张印有一个符号，共有5种符号。莱因要求实验对象通过心灵感应，将牌上的符号告诉另一个人。实验中，一位名叫A. J. 林茨迈耶的实验对象连续15次猜中答

案。于是，莱因欢欣鼓舞地宣布："连续15次准确猜出牌上符号的概率是（1/5）^{15}，约为300亿分之一。"

在这个实验中，"摘樱桃"的错误很难被人发现，错误的原因在于莱因在多轮实验中选择了一轮。如果他真的只做了一轮实验并且实验对象连续猜对了15张牌，他说的概率从技术上讲就是正确的，但是只做一轮实验还不足以保证结果的有效性。也就是说，除了这轮成绩优秀的实验以外，他还做了多轮实验。连续猜对15张牌的那轮测试并不是随机抽取的，而是因为它产生了这种优秀的结果才被选中了，这种行为本身就是在"摘樱桃"。

除了"摘樱桃"，应用不恰当统计方法的错误也比比皆是，这是因为统计方法有很多种，但并不是所有方法都适用于所有情况。最常见的问题也许是样本过小和样本选取不当。很多"软科学"实验的参与者比较少，往往很难得出明确的结论。样本选取不当的问题之所以经常发生，原因是人们在选择参与者时往往会选择有利于某种观点的人。曾经有人在调查最受欢迎汽车的活动中，选择的样本都是当时拥有某个特定车型汽车的人，这是一个非常典型的样本选取不当的例子，这个样本肯定不能代表所有人。

实验设计也有可能影响结果的公正性。在心理学的某些领域，实验者往往希望实验结果与他们的预期之间只有非常细微的偏差。假设在完全随机的条件下，我们预期某个实验得到A、B两种结果的机会各占一半。于是，我们准备进行多轮实验，以便得到更准确的数据。比如，进行25轮实验。选择做奇数轮实验，两种结果就不可能恰好各占50%的比例。

从很多实验可以看出，实验者收集的数据很可能是机缘巧合的结果。经常有报告宣称某种超心理能力肯定存在，因为碰巧出现这些结果的概率非常低。但是，实验者在发布这些信息的时候有些操之过

急。原因之一在于，心理学家预防巧合发生的力度远小于物理学家。如果出现随机条件下发生概率为5%的结果，心理学实验人员通常就会认为这不是一种随机结果，尽管这种小概率结果经常出现。更糟糕的是，仅仅证明实验结果可能不是随机结果，往往不能证明某个假设是真实的，也就是说，不能证明这些是超心理能力作用的结果。

虽然研究人员使用统计工具的过程无可挑剔，但是由于数据解读的问题，他们仍然很难解释实验结果的含义。欧洲核子研究中心大型强子对撞机寻找希格斯玻色子的项目显然就遇到了这个难题。希格斯玻色子是粒子物理学标准模型预言的一种粒子，这种粒子可以使其他粒子具有质量。发现希格斯玻色子与在野外发现一只稀有的老虎是不同的。看到老虎，你可以捕捉、拍照或者取血样并验DNA（脱氧核糖核酸），以确定你看到的确实是一只老虎。但是在寻找希格斯玻色子时，这些方法全部失去了作用。别的不说，实验并没有真的让我们看到希格斯玻色子，而是其他粒子留下来的间接痕迹，实验者认为这些痕迹是希格斯玻色子衰变造成的。由于不是直接证明，实验人员只能通过发生概率来解释，问题也随之而来了。

科学家经常用“西格玛”（σ）这个符号表示标准差这个统计量度。如果把某个事件随机产生某些结果的频率绘制成图，往往就会得到一种叫作正态分布的钟形曲线。比如，手机的重量大多位于某个范围之内，均匀地分布在平均重量的周围。

并不是所有的随机事件都遵循正态分布的规律，因为根据这些信息绘制成的图形不是钟形曲线。有的教科书以人的身高为例，介绍数据集的正态分布特点。但是，这个例子并不恰当。美国男性公民的平均身高大约是5英尺6英寸[注]，从这个数字就能看出一些问题，因为典型美国男性公民的身高（用统计学术语来表示，就是中位身高）超过这个高度。分布图的右侧表示身高越来越高，但在超过平均身高1英尺之后，人数就变得非常少了，超过6英尺6英寸的人更是寥寥无几。然

而，分布图向左延伸的幅度较大，最左侧的身高比平均身高低2英尺多。这个图像并不是真正的正态分布曲线，而是向右“倾斜”，在左侧留下了一个扁扁的长尾巴。

标准差是分布形态的一个量度（只在正态分布这种对称分布中才有效）。标准差表示数据的离散程度，可以告诉我们数据是四处分散还是聚拢在平均数周围。如果数据的离散程度是一个标准差，则表示作为随机事件，统计结果有略高于68%的可能性会落在距离平均值一个标准差的范围内。如果数据的离散程度是两个标准差，统计结果有约95%的可能性落在距离平均值两个标准差的范围内。心理学等“软科学”经常采用这种统计方法。但希格斯玻色子数据分布的离散程度是5个标准差。也就是说，我们所寻找的事件落在距离平均值5个标准差范围之外的概率是350万分之一。但是，如果从他们发现的就是希格斯玻色子的置信度这个角度来考虑，又该如何解释这个实验结果呢?

因此，媒体在报道这项研究时，不得不面对一个可怕的雷区。数据表明大型粒子对撞机给出的结果是一个巧合的可能性非常低。但是，与萨莉·克拉克案一样，我们也不能反过来说，因为巧合发生的可能性非常低，所以希格斯玻色子存在的可能性非常高。数据并不能证明希格斯玻色子可能存在，而只能表明这些数据事出有因的可能性非常高，而且我们猜测造成这个结果的“因”可能就是希格斯玻色子。

更糟糕的是，两者之间的区别十分微妙，几乎不可避免地会造成误读。有的新闻媒体报道，实验结果表明，不存在希格斯玻色子的可能性是350万分之一。但是，统计数据实际上表明，这些数据事出无因的可能性是350万分之一。这项统计指标并没有说实验结果是巧合导致的可能性非常低，而是说在没有原因的情况下产生这些数据的可能性非常低。这就好比一个人说“从这些结果看，事出无因的可能性非常小”（错误），另一个人说“考虑到这是一个百分之百的随机事件，出

现这些结果的频率非常低”（正确），两个人的说法是不一样的。所强调的内容有微小的不同，对于科研的意义却相距甚远。

一言以蔽之，只要运用得当，概率与统计学可以和现实世界实现完美的契合。这样说是有道理的。我们不是利用抽象的数学为现实世界的某个过程建模，而是测量现实世界的某个基于数据的事实或准事实（例如，“抛一枚质地均匀的硬币，得到正面和反面的概率都是1/2”），并在确认这个数据事实成立之后才使用相关的计算方法。与其说我们利用数学探索宇宙的奥秘，不如说我们是在使用数学研究数字的秘密。

即使在使用概率和统计学这两大武器时没有犯错误，我们也会遇到一些问题，主要是因为我们无法轻而易举地洞悉一切。我们通过规律去认识、了解周围的世界，即使有的时候根本不存在任何规律，我们也能“找到”规律。尽管我们知道事件的随机性与非正态分布是它们的真实属性，但我们却感到不舒服。正因为如此，即使专业人士在使用基于概率的统计工具时，也必须小心谨慎。

实践证明，对于以数学为基础的物理学（不仅仅是寻找希格斯玻色子）而言，概率与统计的重要性在不断增加。但是，人们还没来得及证明概率是构成所有物质的粒子的核心属性之一，数学就已经把科学思维推到了另外一个临界点，一个光芒四射的临界点。

---

1. 在三牌赌皇后游戏中，作弊者用一只手拿着三张牌，然后用另一只手将这三张牌展开。在操作时，要让其他人以为他每次拿的都是最下面那张牌。但是，通过不断练习，作弊者可以从最上面拿牌而不被人发觉，尤其是当这些牌稍稍弯曲时，作弊的效果更好。

2. 1英尺约合30.5厘米，1英寸约合2.5厘米。——译者注

第11章
麦克斯韦：关于电磁波的数学方程组

Are Numbers Real?

几年前，我在伦敦的英国皇家研究院（RI）参加过一次辩论会。辩论会的主题很有意思——谁是历史上的第一位科学家。4个人选都不出意外地在本书中占有一席之地。在辩论过程中，阿基米德和罗吉尔·培根（由我提名）的早期成果在伽利略面前相形见绌，后者凭借现代科学成就最终胜出。但是，研究院常驻科学史学家提名的候选人却是比伽利略在历史上出现的时间晚得多的詹姆斯·克拉克·麦克斯韦。

麦克斯韦角逐这个头衔有一个不是很公平的理由——“科学家”这个词在他那个时代才正式出现。在那之前，人们普遍使用的是一个拗口的名称——“自然哲学家”。1834年，人们认为，既然从事艺术工作的人被称作艺术家（artist），那么把从事科学研究的人称作科学家（scientist）似乎是合情合理的。（当时，他们提出了几个备选方案，值得庆幸的是，他们最终没有选择“博学之士”这个词。）但是，在那次辩论会上，人们支持麦克斯韦的理由却更加微妙：麦克斯韦是科学领域中让数学彻底摆脱现实的束缚，并在他提出的理论中有所体现的第一人。

当然，麦克斯韦不是第一个使用数学工具的科学家。我们知道，牛顿早已用他的数学知识构造出一个蔚为壮观的科学魔法世界。麦克斯韦在电磁学领域取得的研究成果虽然与光的本质有关，但是与数学的联系更加紧密。最终版本的麦克斯韦方程组浑然天成，充满美感，完全摆脱了与物理现实的联系，是直接源自数学公式的产物。同时，它也开启了离经叛道的痛苦历程。牛顿处心积虑地为自己的基础数学罩上了一层晦涩难懂的外衣，一旦脱下这层外衣，我们就会发现它特别简单。但是，漫不经心的观察者是无法理解麦克斯韦的研究成果的，他们唯一的选择就是毫不犹豫地接受它。这一点对于科学本身，对于帮助科学赢得社会支持，都具有非常深远的意义。

如果你从未听说过麦克斯韦，也不会令人感到特别奇怪。如果在科学家当中做一次问卷调查，请他们提供三个名字：历史上最伟大的物理学家、有史以来最受他们喜爱的物理学家和最被普通大众低估的物理学家，结果一定非常有意思。艾萨克·牛顿和阿尔伯特·爱因斯坦的名字肯定会出现在第一个名单中，理查德·费曼应该可以轻松地在第二个名单上名列前茅，而詹姆斯·克拉克·麦克斯韦则很有可能成功当选最被普通大众低估的物理学家。值得注意的是，爱因斯坦的书房墙壁上挂着三个人的照片：牛顿、法拉第和麦克斯韦。

介绍麦克斯韦生平的科普读物非常多，但是我认为仍然有必要告诉你们他这个不同寻常的姓名是怎么来的。麦克斯韦的父亲本来名叫约翰·克拉克，但是约翰的父亲继承了克拉克家族麦克斯韦系的庄园和头衔。约翰的父亲死后，约翰的哥哥又继承了这个头衔和位于米德尔比的主体庄园，约翰则继承了位于哥伦莱尔的子庄园，并且从此以后，他在姓名中添加了“麦克斯韦”，以强调这种关系。因此，詹姆斯出生后，他的姓就是克拉克·麦克斯韦，简写为麦克斯韦。

由于家境殷实，麦克斯韦从小就自由自在，而且有机会深入探索附近村庄的自然史。他后来进入爱丁堡大学，攻读物理学学士学位，

之后又进入著名的剑桥大学。毕业后，他先后在剑桥大学、阿伯丁大学和伦敦大学国王学院任职。接着，他离开大学校园，回到哥伦莱尔庄园，潜心做研究。这种没有压力的生活让他感觉十分惬意。直到剑桥大学成立卡文迪什实验室（在这个新实验室的推动下，剑桥大学一举抢占了当代物理学研究的核心地位），冲着第一个卡文迪什教授的头衔，麦克斯韦才又回到剑桥大学，继续他的学术生涯。

在这段时间里，麦克斯韦的研究涉及诸多领域，主要包括运用统计工具从事热力学研究、利用偏振光检测透明材料的应变、探索颜色感知的本质，他甚至还拍摄出有史以来的第一张彩色照片。但是，他留给我们的最有价值的遗产是电磁学。麦克斯韦的前辈、电磁学大师迈克尔·法拉第在思考电磁学领域的各种现象时，曾经提出一个开创性的设想：电与磁会形成一种场。他提出的场概念与等高线图比较相似，相当于整个空间遍布了等高线，等高线的值表示电磁强度。如果一条导线穿过磁场的等高线，就会与磁场发生相互作用，进而产生电流。

法拉第毫无疑问是一位想象力丰富的科学家，但是他没有接受过专门的数学训练，在研究中几乎不会使用数学工具。也有一些科学家曾经用数字来表示电磁感应，但是，就像牛顿对万有引力的理解一样，他们也认为电磁力是一种间接的“超距力”。麦克斯韦无疑是第一个认识到法拉第的场概念的重要意义的人，并且他开始运用数学工具来研究电磁场的概念。从此以后，人类对电与磁的研究发生了翻天覆地的变化。

由于电磁场在三维空间中能无限延伸，在建立数学模型时需要计算空间各点电磁感应的总和。要解决这个数学问题很容易，牛顿与莱布尼茨的微积分恰好可以做到（尽管很少用于三维空间）。但是，麦克斯韦还需要在此基础上再向前一步。按照法拉第的理解，在电和磁形成的“力场”空间中，每个点的值就是磁体或者电荷产生的作用力。

与质量是标量（仅有数量大小）不同，力是矢量，不仅有大小，而且有方向。质量一成不变，而力则必须有确定的作用方向。矢量是由两个数量综合而成的双重数值，在表示力时，它不仅描述了力的大小，还说明了力的作用方向。

在麦克斯韦潜心钻研电磁学的时候，人们刚刚发现了与矢量相对应的数学工具。发明这个工具显而易见很有必要，因为在处理一个矢量时，我们肯定不能把它视为一个简单的数字。比如，两个矢量相加，与其说是代数运算，不如说是几何操作。计算两个矢量之和的最简单方法就是画两个箭头，用箭头的长度表示矢量（例如力）的大小，用箭头的方向表示矢量的作用方向。我们先画一个箭头表示第一个矢量，然后以这个箭头的头部为起始点，画出第二个箭头。从第一个箭头的起始点至第二个箭头的头部画出第三个箭头，这个箭头就是两个矢量的和。但是，麦克斯韦需要解决的问题不只是计算两个不变量的和。在开始研究电磁学之后，他还必须尝试利用向量微积分，计算出空间各点因位置不同而变化不定的矢量。

当时，人们用力学世界观研究世间万物，麦克斯韦刚开始的时候也没有彻底放弃这个工具。他考虑过将电磁力（电与磁的综合体）看作并不存在的多孔固体中的流体，用流体代表电磁力的力线。同所有优秀的模型一样，他的这个想法不仅与建模时观察到的各项数值相互吻合，而且用模型做出的一些预测也得到了证实，其中包括电磁力的强度与距离的平方成反比。

麦克斯韦的“渗流”模型是一个非常有用的工具，但是它的作用也仅限于此。重要的是，代表力线的流体被固定在固体内部的通道中，但是在绝大多数情况下，例如，在越来越受欢迎的法拉第电动机和发电机中，力线并不是静止不动的。要表示这种情况，麦克斯韦必须彻底改变自己的模型，但他暂时无能为力，原因之一就是他在这前后收拾行李搬到了伦敦。5年之后，他才开始考虑这个问题。

在这种情况下，平庸的科学家很可能会修改自己的模型，以适应运动的电磁力。但是高明的科学家在意识到错误时，即使他们在这条道路上已经投入大量的时间和精力，他们也会当机立断，另辟蹊径。麦克斯韦果断地放弃了他的流体模型（即使是现在，运动的流体也是最难处理的物理学问题之一，很难得到精确的结果），转而到更熟悉的力学领域中寻求答案。他首先对磁体进行了研究，发现他的模型需要添加磁力线，而且磁力线上应该有朝着某个方向的应力（与异性磁极之间的吸引力相对应）和垂直压力（与同性磁极之间的排斥力相对应）。

他设想，磁体可能是由一系列可以自由旋转的微小电池构成的。地球等实体旋转时，在旋转力的作用下，赤道附近的地方会凸起，两极则会稍扁。与此同时，地球内部的微小电池也会发生同样的变化。如果一系列电池的旋转轴相同，在这些电池被压扁的时候就会在沿着旋转轴的方向上产生应力，而在微小电池凸起的地方就会产生向外的垂直压力。这种效应与他的电磁场模型正好一致。

至此，他的模型没有任何问题，但是在实际操作时，这些微小电池常常因为相互作用而僵持不下。原来，麦克斯韦忽略了电荷的影响作用。为了解决这个问题，他设想每个电池周围都有一些很小的球体，就像安装在转轴周围用来减小摩擦力的滚珠轴承一样。微小电池只能在固定位置旋转，而那些小球体却像电流一样，可以在原材料中穿行。尽管麦克斯韦的这个模型十分简陋，但令人吃惊的是，它与金属由原子构成、电子从原子中间穿行的结构非常接近。直到多年之后，人们才发现原子存在的证据（又过了更长时间才知道电子的存在）。

此时，麦克斯韦的模型已经可以解释某些电磁现象了，但是它无法解释电磁感应——这个物理现象对于变压器具有非常重要的意义。一条电线中的电流发生变化，为什么会导致另一条电线产生浪涌电流

呢？法拉第正确地推断出电流在接通和断开时会产生磁场，而磁场可以通过另一条电线扩展、收缩，进而产生电流。麦克斯韦利用他的微小电池（此时他已经把这些电池改成六面体了）和小球体，成功地模拟了这个变化过程。为了模拟电磁感应现象，他对模型做了一些改进。当小球体从微小电池周围流过时，不同层次的电池会做出不同的反应，同时会对小球体的运动施加阻力，使这些小球体的速度逐渐减慢。

做了这些改进之后，他的模型已经比较完美了，唯一的缺陷就是不能很好地表现电荷之间的相互作用。梳理过干燥头发的梳子带有电荷，可以吸住碎纸屑，原因就在于电荷的相互作用。同很多前辈一样，麦克斯韦也发现遇到问题时暂且放一放，反而更容易找到解决办法。

他想到，纸、瓷器等绝缘体内部的小球体被固定在微小电池上，不能像金属内部的小球体那样自由流动。他又想到，如果绝缘体的微小电池具有弹性（在固定位置上可以发生扭曲），这些电池就可以像弹簧一样，通过扭曲将能量暂时储存起来。与之相比，金属的微小电池是刚性的，几乎不会发生扭曲。这个想法不仅说得通，它还可以高度精确地模拟真实材料的特性，从而帮助我们了解电磁力的真实属性。根据这个模型，静电力与弹簧的势能非常相似，磁力则更像转动能，而且两者一定会产生相互作用，不可能独自出现。所有问题似乎都得到了妥善解决。但是，就在此时，麦克斯韦却发现了一个足以让整个模型变得一文不值的大问题。

法拉第场被认为无处不在，甚至存在于真空中。由于空间具有极强的绝缘性，因此，根据他的模型，空间应该包含弹性电池。弹性物体的一个特点是它可以传递波。事实上，弹性是波传播的必备条件。由此可见，电磁波似乎可以在真空中穿行。此外，弹性电池的扭曲会产生磁场，磁场又会拉扯邻近的小球体，进而形成电场。麦克斯韦认

为，真空中应该有一个包含电波和磁波的自持波，其中电波和磁波彼此垂直，而且都与自持波的传播方向垂直。同时，电波会产生磁波，磁波又会产生电波，循环往复。

当时，人们已经确认光是一种独特的波，在介质中传播时会发生左右摆动。（通常，只有在介质边缘传播时，波才会左右摆动。）人们还通过实验发现，光与磁之间似乎存在某种联系。麦克斯韦在想，根据他的模型，光就是一种电磁波，这个近乎荒谬的结论到底是不是真的呢？毕竟，光可以毫不费力地穿透真空，从太阳传播到地球。随后，麦克斯韦对这种假设的波的传播速度进行了估算，结果发现它与光在真空中的传播速度非常接近。这个发现进一步展现了这个模型的惊人效力。

毫无疑问，麦克斯韦在数学工具的使用方面远远超过他的前辈，这也是他当时最令人瞩目的成就。但是，在对气体黏滞性进行了初步研究之后，他又一次放弃了自己建立的模型，然后重新开始研究电磁学。这一次，他没有使用流体、旋转式弹性电池等类比模型，而是采用了现代物理学家都非常熟悉的方法，即建立纯粹的没有掺杂其他任何内容的数学模型。

虽然他使用的仍然是建立模型的方法，但这个模型仅仅是一些表现一系列逻辑法则的数字。模型中没有图，没有类比，也无法让人们轻松掌握其中的原理。麦克斯韦使用的是18世纪意大利数学家约瑟夫–路易斯·拉格朗日发明的方法——拉格朗日函数。借助拉格朗日函数，人们可以通过微分方程组，以及系统各部分的动量、系统动能等要素，描述系统随时间的变化而发生变化的情况。

拉格朗日函数就像一个黑箱（在数学家的眼中极具简洁雅致的美感），只需输入已知因素，摇动把手，就能得到答案。使用者根本不需要深入了解系统的物理属性，因为建模时使用的全部是数字。

为了满足电磁学的特殊要求，麦克斯韦进行了一番复杂的改进，把拉格朗日的研究成果变成了一个比较简单的方程组，用来描述电与磁的特性。麦克斯韦的后辈们利用现代符号表示法，对麦克斯韦方程组进行了完善和精简，最后得到了4个令人惊叹的简短方程：

$$\nabla \cdot D = \rho_f$$

$$\nabla \cdot B = 0$$

$$\nabla \times E = -\frac{\partial B}{\partial t}$$

$$\nabla \times H = J_f + \frac{\partial D}{\partial t}$$

对于现代物理学家而言，这些方程简单明了、司空见惯（然而大多数人却会惊叹于它们的复杂程度）。但是，在麦克斯韦提出这些方程组时，大多数科学家都觉得，单凭数学工具就构建出这个模型，实在是一件不可思议的事情。如此彻底地摆脱对现实世界的依赖，是很多人努力追求的目标，其中包括与麦克斯韦同时代的杰出人才、在更年轻时就已成为物理学教授的威廉·汤姆森（即开尔文勋爵）。由于没有人亲眼见过电磁波，因此人们对这个模型的反响并不强烈。但是，这套理论认为，只需让电荷沿着一片金属（现在的人称之为天线）运动，就可能产生电磁波。直到20年后，海因里希·赫兹用这种方法首次生成无线电波，麦克斯韦的这项突出成就才真正得到人们的普遍认可。

尽管麦克斯韦可以借助数学工具完成他的研究，但在当时，就连他本人也不相信由这些公式得出的所有推论。他喜欢借助数学工具建立模型，但是他认为那些数字不可能直接反映现实世界的本质。麦克斯韦曾经两次无视这些数字给他的提示：一次是他在提出一个假说时没有考虑这些数字；另一次是他直接忽略了这些方程给出的某个预测结果，原因是这个结果太奇怪了。

那个假说与以太有关。麦克斯韦的电磁波理论成立的前提条件是，真空有容留电磁场的能力，因为电磁波离不开电磁场。波的传播不需要任何介质。如果真的存在类似于普通物理材料的介质，这些波就会显得特别奇怪，因为介质会抑制波的左右振动，导致横波无法从介质中间穿过。通常，这样的波都在介质边缘传播（例如，在水面上或者在小提琴琴弦上）。能从介质中间穿过（例如声音在空气中传播）的波往往是压缩波，也就是波的振动方向与传播方向一致。

尽管麦克斯韦的模型明确地告诉他，光从恒星发出后，无须以太也能在真空中穿行，但他仍然坚信以太肯定存在。优秀的科学家几乎都兼具叛逆者和传统主义者的双重特点，他们肩负着破旧立新的使命，但是他们又不可能推翻一切、从头开始。他们必须依赖某些现存的观念，但是这些观念往往已经存在很长时间，以致失去了继续存在的意义。以太就是一个这样的概念。有趣的是，包括诺贝尔奖得主弗兰克·韦尔切克在内的一些现代物理学家认为，以太从一定意义上讲仍然是存在的，但是我们需要换一个思路，把数学意义上充斥整个空间的各种场（例如电磁场）视为以太。

麦克斯韦忽略的那个预测极具震撼性，令人震惊的程度远胜过以太是否存在这一问题。这个预测表明，一定存在可以回到过去的波。为了说明这个问题，我们有必要先花点儿时间，思考一个非常简单的数学问题。在下面这个方程中，x的值是多少呢？

$$x^2 = 4$$

即使“代数学”这个词令你惊恐万分、深恶痛绝，解这样的方程你肯定也会胸有成竹。我们的任务就是找到平方为4的x的值。不难发现，2是满足条件的答案。但是，如果在学校考试中解这个方程，回答2只能得到一半的分数，因为答案不止一个。x等于–2时方程同样成立。也就是说，这个方程有2和–2两个解。

某些方程（例如二次方程，上面这个方程就是一个简单的二次方程）经常会出现有两个解的情况。麦克斯韦方程组在预测电磁波这种自持波存在的可能性时，同样遭遇了这种情况。这些方程的解不是一个，而是两个，并分别被称作“延迟波”和“超前波”。根据这些方程，当我们熟悉的电磁波（包括无线电、X射线、伽马射线在内的所有光）从A传播至B时，这些波都是延迟波。但是，这些方程还描述了第二种波，即从B传播至A的超前波。在延迟波到达B的那一瞬间，超前波从B出发并逆时传播，在延迟波离开A的瞬间到达A。

这个预测显然面临两大难题。难题之一是，任何人都没有见过超前波。如果超前波真的存在，那么它们似乎可以完成时光倒流这种不可能实现的壮举。尽管数学上没有给出任何提示，告诉人们应该只保留其中一个解，忽略另外一个解，但是大家都这样做了，因为他们认为那个解非常奇怪，不应加以考虑。数学工具似乎给出了一个与现实世界格格不入的预测结果，但是，由于这些方程式完美地描述了电与磁的其他特点，所以人们无法弃之不用。

直到20世纪40年代，美国的两名物理学家约翰·惠勒和理查德·费曼才发现超前波不仅是这些方程的一个预测结果，它还可以在物理学领域发挥重要作用。尽管科学的发展在一定程度上需要科学家解放思想，但是既有的科学理论对大多数科学家的阻碍作用仍然十分明显。然而，惠勒和费曼都十分开明，不会被常识遮住双眼。

为了解释光与物质之间的相互作用，费曼等人创立了量子电动力学（QED）这个物理学分支，而走在时间前面的超前波将帮助他们解决一个问题。量子电动力学常常会引出数学意义上的无穷大概念（参见第12章），这个缺点不仅会影响量子电动力学的发展，还会给量子色动力学等现代理论惹来麻烦。当惠勒和费曼提出他们的大胆想法时，量子电动力学面临的电子反冲问题就属于这一类型的麻烦。如果原子中电子的能量降低并释放出一个光子，这个电子就会发生反冲，

这与枪支发射子弹的情形十分相似（光子没有质量，但是它们有动量，而动量一定是守恒的）。

要让电子发生反冲，电子的电场必须作用于电子自身，因此它们实际上构成了一个反馈回路，而且会导致无穷大的结果。但是，当时的人已经知道电子会不停地释放光子，我们看到的光大多就是这样产生的。惠勒和费曼发现，如果每次产生的光子不止一个，而是两个，而且其中一个是逆时运动的超前光子，在解释反冲问题时就不会导致无穷大这种令人无法解释的结果了。

毫无疑问，惠勒和费曼的研究得出了一个有用的结果，但是人们一直认为这是一种效果不错的数学魔术，对于揭示现实世界的本质没有任何启示作用。当然，使用过这个研究方法的人大多（不包含惠勒和费曼）不认为物质世界中真的存在超前波和超前光子，但是这个例子再次说明数字可以产生与实际观察相匹配的结果（尽管本例中的这个结果有点儿出乎人们的意料）。常识告诉我们，波不可能进行逆时传播，但是数学工具却告诉我们相反的预测结果。事实证明，数学工具做出的稀奇古怪的预测，与更加直接的研究方法相比，其反映现实的效果更好。

对于现代数学界而言，无论我们怎么理解麦克斯韦方程组的解，这些方程都没有多大的难度，尽管当初提出这些方程的人是一个天才。然而，正如与麦克斯韦同时代的格奥尔格·康托尔发现的那样，即使是专业的数学人员，也会遇到解决不了的难题。

第12章
康托尔：让一众科学家挠头的无穷大

无穷大这个概念一直令人耿耿于怀，其实并不奇怪。从古希腊时期开始，人们就开始思考无穷大是否存在、本质是什么的问题。古希腊人肯定知道，用于计数的正整数序列没有尽头。如果真的有最大的整数（我们用*max*来表示这个数），就必然有$max + 1$、$max + 2$等更大的数。但是，无穷大概念让古希腊人很不舒服，他们用来表示这个概念的“阿派朗”（apeiron）一词就有混乱的含义。

哲学家亚里士多德就曾对这个概念进行过研究，他的观点在随后几百年时间里一直占据主导地位。公元前384年，亚里士多德出生于希腊北部。他认为，无穷大具有必然性，但是又无法达到。他从世间万物中找到了一些他认为属于无穷的例子，例如，他认为整数（我们已经知道整数是无穷的）和时间就是无穷的。此外，他还认为有的东西是无限可分的。但是，与此同时，他又提出了几个含混不清的观点，想证明无穷大不可能存在于现实世界中。例如，他说任何物体都有边界，如果某个物体是无穷大的，它就不可能有边界，也就不可能存在。

在明显经过一番艰苦的心理斗争之后，亚里士多德最终断定无穷大不是一种存在于现实世界的概念，而是一种潜在的可能性，在现实中永远无法实现。无穷大是存在的，但是不能根据需要变为现实。他以古代奥运会比赛为例，简单明了地介绍了自己的这个想法。比赛毫无疑问是存在的，肯定不是一个虚构的概念。但是，一般而言，如果有人请你把奥运会比赛展示给他看，你肯定做不到。因此，奥运会比赛就是一个潜在的实体，你没有办法仔细端详，小心鉴别。亚里士多德小心翼翼地指出，尽管某些潜在实体在特定时间或特定空间里会变成现实，但是无穷大不包括在内。

牛顿和莱布尼茨创立微积分的时候，使用的正是潜无穷的概念（参见第9章）。微积分中的无穷大是一种极限。我们非常熟悉的双纽线符号（∞），表示的就是这种可望而不可即的目标，也就是亚里士多德所讲的潜无穷。与牛顿同时代的约翰·沃利斯在一篇枯燥乏味的论文中第一次使用了这种符号。但是，沃利斯仅仅说了一句“用∞表示无穷大”，却没有解释这个符号从何而来。

直到19世纪，绝大多数数学家都认为亚里士多德的观点是正确的。他们普遍认为，潜无穷这个概念是正确理解无穷大的唯一可行的途径。例如，享有盛名的19世纪德国数学家卡尔·弗里德里希·高斯明确地说：

> 我反对将无穷量视为一个实体，这在数学中从来都是不允许的。所谓无穷，只是一种说话的方式，其真正的意义是指某些比值无限接近于某个极限，而另一些比值则可以无限增大。

但是，并不是所有数学家都被这种思想蒙蔽了双眼，伽利略就是这些异类中最引人注目的一个。一说到伽利略，我们首先想到的是他因为支持哥白尼的日心说而受到宗教裁判所的审判并被终身监禁。但是，伽利略对科学领域做出的最重要的贡献其实是他于1638年出版的

《两种新科学的对话》。这是他在物理学领域的代表作，为牛顿成功完成关于力学和运动的研究奠定了基础。

同他支持哥白尼的日心说、给他带来无数麻烦的那部著作一样，这部新作也采取了三人对话的形式（这在当时十分流行）。它放弃了古板的拉丁语，改用意大利语，可读性远胜牛顿用专业术语撰写的几乎无法理解的作品。在此之前，伽利略因为出版著作已经被判终身监禁，这本书可以成功出版的确不是一件易事。最初，伽利略准备在威尼斯出版这本书。当时，威尼斯因为从罗马帝国独立出去而自视甚高。但是宗教裁判所发出了一系列禁令，禁止出版伽利略的所有作品。伽利略只有获得宗教裁判所的批准，才能出版自己的这部新作。

顽强不屈是伽利略最显著的特点。尽管受到禁令的限制，而且间接违反禁令的钻空子行为将会为他带来危险，但伽利略没有放弃。1636年，荷兰出版商洛德韦克·埃尔泽维尔造访意大利，伽利略想办法把手稿送给他看。在最终付梓成书时，人们发现它的献词非常有意思。之前，伽利略每次都会把他的作品献给权势人物，而且从不吝啬他的赞美之词（很可能是因为当时谄媚成风）。作为交换，他有可能得到对方的资助。但是，这本书却被献给了他以前的学生、法国驻罗马大使弗朗索瓦·德·诺瓦耶，而且他在写献词时十分小心。显然，他绝不希望给诺瓦耶惹来麻烦。

从字里行间可以看出伽利略既有简单直接的一面，又有迂回曲折的一面。宗教裁判所受他蒙骗的可能性非常小，但他们似乎并不是很重视这件事。伽利略称：

> 我已经决定不再出版任何成果。但是，为了我的研究不被彻底遗忘，我觉得有必要在某个地方留一份手稿，至少让那些研究相同内容而且方法得当的人有机会看到它。因此，我首先想到要将我的研究成果交到阁下手中……

伽利略既感谢诺瓦耶的热心帮助，又不想让人们认为诺瓦耶是负责这本书的出版事宜的人，因此，他抛出了几个神秘的中间人：

> 埃尔泽维尔告诉我他们准备出版我的这些成果，要求我起草献词并且立刻做出答复。这些出版商曾经出版过我的一些成果，因此他们希望继续出版我的这项成果，并把它变成漂亮的精装版书籍。这个消息让我深感荣幸，也有些措手不及。我知道，他们能拿到我的手稿，是因为阁下为了帮助我恢复和传播名声而把我的成果介绍给了几位好友。

他对诺瓦耶心怀感激，同时又表示，手稿到达出版商的手里是诺瓦耶的几位不知名的朋友促成的。很显然，伽利略在这本书即将付印时才收到消息的说法是不真实的。早在埃尔泽维尔造访意大利时，伽利略就想尽办法把书稿送到了他手上。不仅如此，他和埃尔泽维尔还有书信往来，就这本书的内容进行过深入探讨。伽利略是一位令出版商咬牙切齿、捶胸顿足的作者，因为不到最终出版他都不会停止修改书稿。即使在采用电子印刷技术的现代，这种行为也会导致相当大的麻烦。而在当时，人们必须利用活字印刷术小心翼翼地将所有书页制成一个个印刷版，因此，任何修改都是一场可怕的噩梦。然而，无论是因为宗教裁判所遭受了蒙蔽，还是因为他们睁一只眼闭一只眼，这本书最后还是成功出版了，但伽利略的祖国意大利并没有公开发售。

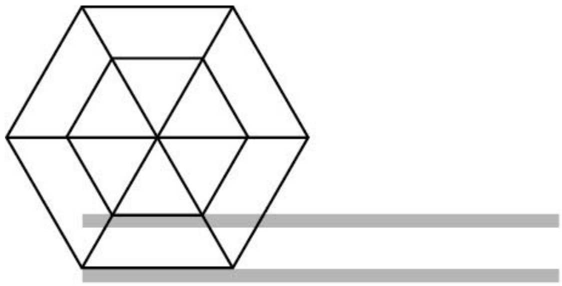
书名中所说的“两门新科学”是指对固体物质的本质研究与对运动的分析，无穷大的概念出现在这本书的第一部分。在分析固体物质紧密结合在一起的原因（例如，金属块为什么难以分割）时，书中的一位主角认为，构成这些固体物质的粒子之间的真空将它们紧密地吸附在一起。（真正的原因其实是电磁作用，因此他的这个观点是不正确的，但听起来有些道理。）这个说法遭到了辛普利西奥（书中三个人物之一）的质疑，辛普利西奥是古希腊思想的信徒，他的任务就是质

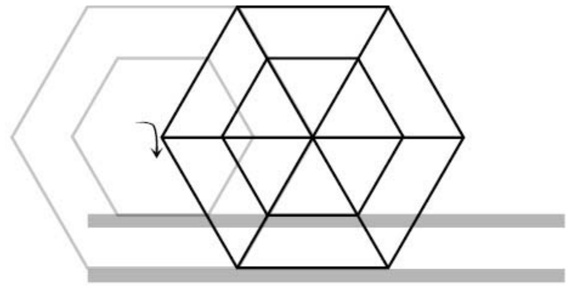
疑新思想。他认为，由于空间十分狭小，即使有真空存在，作用力也会非常小，远不足以让金属紧密地吸附在一起。

于是，三个人继续思考。很多微弱的作用力加到一起，是不是可以变成一股异常强大的作用力呢？三人中的萨格雷多说：“也就是说，数量庞大的蚂蚁群可以抬起装满谷物的船。”他认为，无数个微小的作用力加在一起，就可以克服任何阻力，只要这个阻力不是无穷大的。萨格雷多的“只要这个阻力不是无穷大的”这个条件，遭到了萨尔维亚蒂的嘲笑。萨尔维亚蒂在这本书中主要是扮演伽利略的代言人，他说，即使一个物体的大小有限，其内部也完全有可能存在无穷多个真空。

这似乎是一种托词，伽利略的真正目的可能是提出几个与无穷大有关的有趣想法，因为他接下来又用大量篇幅探讨了无穷大的本质。与亚里士多德提出的那个说服力不足却令数学界认可了很多年的潜无穷概念不同，伽利略给出的是一个实实在在、没有任何掩饰的事实。他画了一个想象出来的奇怪的组合图形，用来说明无穷大的神奇特性。

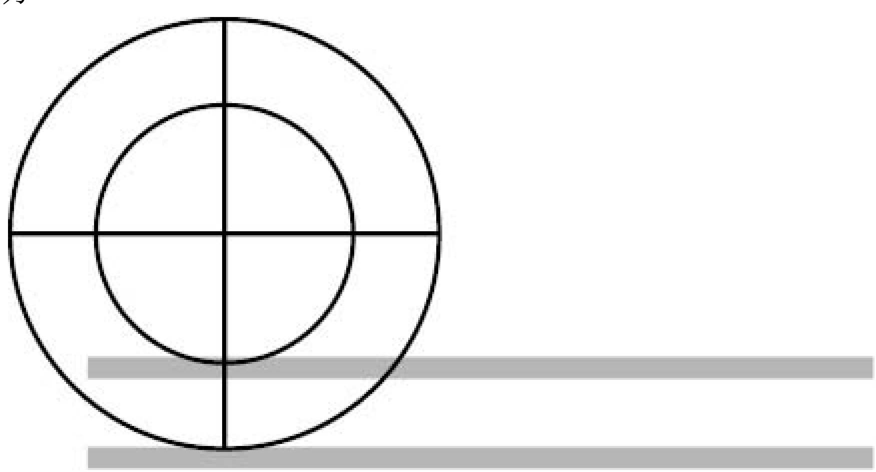
这个组合图形是由大小不同的两个六边形组成的，小的六边形贴在大的六边形前面，6个角对齐，都位于水平轨道上。然后，萨尔维亚蒂请另外两个人想象着把这两个六边形转动1/6圈。这个六边形“车轮”向前移动的距离应该等于大六边形的边长，因为现在是第二条边在最下面。这个结果没有什么稀奇的地方。但是，我们知道小六边形的边长要短得多。小六边形沿着轨道转动了1/6圈，因此它移动的距离应该是小六边形的边长，但是实际上，它前进的距离却等于大六边形的边长。





这个现象并不难解释。大六边形转动时，小六边形就会从轨道上被抬起，向前跳跃，跳跃的距离正好等于这两个六边形边长的差。因此，小六边形不仅向前移动了一个边长的距离，还向前跳跃了一段距离，两者之和正好等于大六边形的边长。

到目前为止，这个解释没有任何问题。接着，伽利略想，如果多边形的边数不断增加，结果会怎么样呢？随着边数增加，车轮转动1/6圈时，大多边形就会有更多的边参与这个过程，同时，小多边形需要完成更多次的小幅跳跃。在接下来的想象中，伽利略展现了他的聪明才智。他不断增加多边形的边数，直至车轮变成圆形。此时，多边形的边数实际上变成了无穷大。在这种情况下，如果整个车轮转动1/6圈，小车轮也会转动1/6圈，但是它仍然可以与大车轮保持同步，向前移动相同的距离。此时，由于车轮是标准的圆形，因此它应该不会在轨道上跳动。



这就令人感到迷惑不解了（辛普利西奥对此感觉尤为奇怪）。伽利略借萨尔维亚蒂的口说，这是因为小车轮完成了无穷多个幅度无穷小的跳跃，这些跳跃的总长度，正好弥补了小车轮移动距离上的不足。萨尔维亚蒂一面满怀愧疚地承认这个事实令人震惊，一面又请求不妨放下该书当前讨论的内容，转而对无穷大这个概念加以研究。另外两个人也高兴地同意了他的请求。

他们先举了一个晦涩难懂的例子，并用几何方法证明了点的集合与圆的圆周有可能大小相同，然后又回过头来继续讨论这些车轮。辛普利西奥发现，第一个例子似乎包含两个无穷大：大车轮周长的1/6是由无穷多个点构成的，小车轮周长的1/6也是由无穷多个点构成的。这两个数都是无穷大，但是一个无穷大对应的结果却大于另一个。萨尔维亚蒂先是敷衍搪塞，说从有穷的角度去理解无穷的概念，就会导致这个问题，但紧接着他又试图向辛普利西奥证明，这是无穷大的内在属性造成的一种奇怪现象。

他的证明使用了正整数（即自然数）的平方数这个概念。萨尔维亚蒂说，每个自然数都对应一个平方数。辛普利西奥欣然同意这个说法。接着，萨尔维亚蒂又问他，自然数有无穷多个，而且每个自然数都对应一个平方数，那么这些平方数的个数是不是等于正整数的个数呢？答案显然是肯定的。但是，正整数中还有很多数本身并不是平方数，例如2、3、5、6、7等。也就是说，每个平方数都会对应一个自然数，而自然数的个数远多于平方数。

伽利略通过这番讨论明确地告诉我们，传统的运算法则不适用于处理实无穷。此时，“相等”、“小于”、“大于”等概念也会失去它们的传统含义。我们可以说一个无穷集（例如正整数集）可能包含无穷子集（例如平方数集）。伽利略笔下的这三个人之所以遭遇麻烦，原因之一就是他们把无穷大看作一个数字（伽利略就是这样想的）。我们现在不会把无穷大看作一个数字。我们可以说某些事物构成了一个无

穷集，但不会说这是一个无穷大的数字。如果伽利略晚出生200年，就会明白其中的道理。

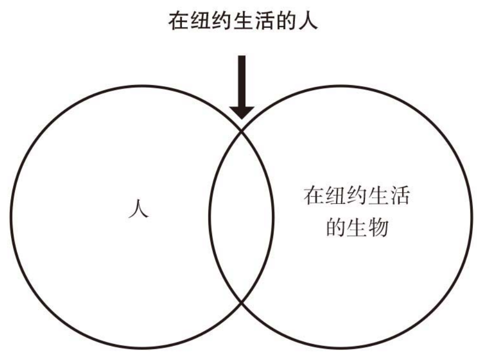
在伽利略之后，所有人都把眼光投向了更容易让人接受的潜无穷概念。直到19世纪，格奥尔格·康托尔决心揭开其中的真相，才改变了这种状况。康托尔是一名数学家，但是由于他坚信无穷大是一种真实的存在，再加上其他数学家都认为他是在引火烧身，所以，不仅他的职业生涯充满了艰辛，他的精神世界最终也轰然倒塌。康托尔认为，数学和数学家都可以接受无穷大真实存在这个赤裸裸的事实。他试图证明有比无穷大还大的存在。这项似乎根本不可能取得成功的研究，在现实世界中没有明显的实用价值，但康托尔取得的成就并不只是这些。

康托尔还是集合论的创立人，集合论似乎可以解释数学的作用原理。要感受康托尔在无穷大研究领域中表现出来的杰出天赋，我们有必要对集合论稍加了解。在本书第1章，我提到了数字到底是什么以及它们与周围世界存在什么关系的问题。集合论对数字进行了形式上的定义，而且这个定义显然是以现实为基础的，但它又摆脱了现实的束缚，卓尔不群地屹立在柏拉图洞穴外面的数学世界之中。集合论对于数学的意义就相当于原子论对于科学的意义。我们曾经懵懂无知地生活了几千年，但是在接受了原子的存在之后，我们就认同它们是构成自然界的基本单位。同样，在几千年的时间里，数学研究并没有因为集合论的缺失而令我们感到任何不适。但是，集合论问世后就立刻变成一切数学研究的基础。

所谓集合，是指一系列具体事物或者抽象概念。这些事物或者概念都有一个共同的特点（例如一组名叫“布赖恩”的事物或者一组看上去像甜甜圈的事物），或者是基于地点或时间建立起某种联系的随机组合（例如纽约人行道上的所有事物，或者你今天上午想起来的事情）。集合论的某些表达还进入了人们的日常用语。“子集”是指在一

个集合中拥有另外某个共同特点的所有元素的集合，是包含在一个大集合中的集合。例如，“美国人”这个集合是“人”这个大集合的一个子集。集合中的各项称作该集合的元素，也就是说，只要你不是智能机器，你就是“人”这个集合中的一个元素。

大家可能见过用维恩图这种直观的方法表示的集合。用维恩图来表示集合的相交与合并，这是很容易理解的。例如，我们可以用下图表示“人”与“在纽约生活的生物”这两个集合（除了人以外，后者还包含很多其他元素）相交的情况，其中重叠的部分代表“在纽约生活的人”。



使用搜索引擎时，我们经常会不自觉地使用各种集合。借助“与”“或”“非”等布尔代数术语，我们可以进行集合的合并或者选择。例如，如果你使用下面这个搜索项在网上搜索图片：

（车与美国）（福特或雪佛兰）（非红色）

那么你搜索的就是一个子集：美国车，品牌为福特或雪佛兰，除红色以外的其他颜色。至少以前的搜索引擎是这样工作的。现在的搜索引擎（例如谷歌、必应等）都自视甚高，因此大多不屑于使用这些布尔代数术语。

在处理集合时，数字有两种截然不同的用法，使用时一定要注意区分。本书中使用的1、2、3等数字都是“基数”，这是自然数的主要用

法。但是，我们也可以利用数字来规定各子集在集合中所处的位置，此时这些数字叫作“序数”。例如，在考虑橙子集合时，数字3可以指集合中橙子的数量，也可以指集合中的第三个橙子（“3号橙子”）。

我们往往认为序数是数字的一个有用但不怎么重要的用法，但一些人类学家指出，序数的出现早于基数。如果情况属实，我们在第2章里介绍的数山羊活动就应该是另外一种意义了。这些人类学家认为，计数首先不是出现在像贸易这样平淡的事件之中，而是出现在宗教仪式之中，因为宗教仪式中的重要事物必须以正确的次序出现。也就是说，表示次序的数字出现在表示物体数量的数字之前。这些人类学家并没有找到有说服力的证据，而且人们有足够的理由认为，这些人类学家提出这个观点的目的可能是试图说明他们的人生观远比会计人员的人生观更重要。当然，有序计数有可能出现得更早，尽管我们看不出它比掰手指数山羊的方法更有优越性。

描述集合的大小（也称集合的势）非常有用，无论这些集合的大小是否可以表示成数值的形式，我们都可以通过势来比较大小。如果我们设想将两个集合并排，让两个集合的元素结成对，并且形成一一对应的关系（一个集合中的每个元素都能在另一个集合中找到唯一一个与之对应的元素），那么无论我们是否知道这些集合的大小，我们都可以说这两个集合等势。在研究无穷大的概念时，上面说的形成一一对应的关系将发挥非常重要的作用。

如何利用势来比较集合的大小呢？我们可以想象有两个集合，一个是指南针方向构成的集合，另一个是季节集合。我们可以让北与冬季、东与春季、南与夏季、西与秋季分别配对。这样，我们将所有方向和季节都包括进来，而且一个集合中的元素分别与另一个集合中的元素构成一一对应关系。因此，即使我们不知道一共有多少个不同的季节，也不知道有多少个不同的方向，我们也可以说这两个集合等势。事实上，在这个例子中，我们知道集合的元素数量是4。但是，重

要的是我们无须知道这个数字。只要可以重复这个一一对应的程序，我们就会知道这两个集合等势。

请大家回想一下令辛普利西奥感到困惑的那个奇怪的现象。每个自然数都可以与一个平方数配对，因此我们知道，自然数集与平方数集等势。但是，我们还知道平方数构成了自然数集的一个子集。后来，康托尔发现，无穷集一定包含一个与自身等势的子集。

在康托尔开始研究集合之前，意大利数学家朱塞佩·佩亚诺就已经利用集合来定义自然数了。在第2章，我们根据真实物体建立了数字系统，尽管最初是从数山羊开始的，但是这套数字系统最后可以应用于所有物体。佩亚诺摆脱了真实物体，单独考虑这些数字，使它们可以仅凭集合的本质而独立存在。这个方法在早期数学中无法采用的原因之一是，它必须建立在“无”（0的化身）这个基本概念的基础上。具体来说，这个方法的起始点是空集，即不包含任何元素的集合，这为数字0的产生奠定了基础。

第二个集合只包含一个元素——前面定义的那个空集。通过这个方法，佩亚诺得到了基数1。接着，他又创建了一个集合，将前面的那个集合包含其中。也就是说，这个集合包含两个集合：空集和包含空集的那个集合。这样，他又得到了基数2。以此类推，只凭一个个集合，我们就像俄罗斯套娃那样，搭建出“自然数”的完整集合（包括0和正整数）。

物理学家罗杰·彭罗斯认为，既然可以用这种方法定义数字，就说明：“只需发挥想象力，这些数字就会栩栩如生地出现在我们的脑海里，我们可以充满信心地使用它们，而不需要考虑物质世界的任何属性。”然而，我认为这个理由包含了不可靠的诡辩术成分。毫无疑问，我们无法仅凭想象力就让自然数“出现”在我们的脑海里，或者说我们无法完全脱离具体物体，凭空想象出这些数字。

而且，所谓集合，就是一系列实体。要建立集合的概念，首先必须有这些实体存在。如果现实世界中没有可以计数的物体，很难想象我们会产生集合的概念。比如，我们假设世界上存在一种有思维能力的生物，它既没有具体形状，又与物质世界没有联系。既然这个生物除了自己的存在以外，得不到任何其他体验，那么它怎么能像我们一样感知到周围世界的多样性呢？它又怎么可能产生自然数、集合等概念呢？

佩亚诺和康托尔借助集合论，使算术脱离了数山羊的现实活动，为数学中的数字奠定了理论基础。从某种意义上看，这个抽象化过程帮助数字摆脱了计数作用，直达数字的本质，因此具有非常显著的意义。但是，这个变化过程也使某些数学家感到不安（时至今日，他们仍然不能释怀），这是因为集合论的核心理论包含一个令人不安的悖论。第一个指出集合论面临这种困境的人是英国哲学家、数学家伯特兰·罗素。

佩亚诺通过创建以其他集合作为自身元素的集合，推导出了自然数。与之相似，罗素研究的是另一种包含子集的集合，具体地说，他是从集合是不是自身的元素这个角度展开研究的。这个说法似乎会导致棘手的递归问题，但是，通过具体实例，我们就可以洞见其中的奥秘。例如，我们考虑“狗”这个集合。与这个集合对立的是一个更大的集合——“除了狗以外的所有事物”。假设我们认为“所有事物”不仅仅包含具体事物，那么“除了狗以外的所有事物”就是它自身的一个元素，因为它是一个非狗集合。同理可知，“狗”这个集合不是它自身的元素。

接下来，罗素提出了一个新颖的问题：考察“不是自身元素的所有集合”这个集合。这个集合包含“狗”这个集合，但是不包含“除了狗以外的所有事物”这个集合。我们把这个新的集合称作“非元素”集合。罗素问道：“非元素”集合是不是它自身的一个元素？

至此，我们的大脑很可能已经陷入困境而难以自拔了，当年的罗素就遇到了同样的麻烦。如果“非元素”集合是自身的一个元素，那么根据定义，它就不是自身的一个元素，因为这个集合就是这样定义的。同样，如果“非元素”集合不是自身的一个元素，那么它应该是自身的一个元素。从逻辑上讲，“我说的这句话是谎言”这句话与“非元素”集合有相同的效果，都会导致自相矛盾的结果。事实上，罗素要告诉我们的是，集合论有一个固有的内在矛盾，这是数学家无法容忍的。但是，集合论仍然是数字本质和简单算术的基础。

我们将在下文继续讨论罗素发现的集合论问题，但是现在我们先花点儿时间，看一看康托尔的研究，了解无穷集合的定义。如果将佩亚诺构建自然数的方法发挥到极致，就会得到一个无穷集合。这与用双纽线符号表示的潜无穷有所不同，因此康托尔发明了一个新的符号——$\aleph_0$，它是由希伯来文的第一个字母与0构成的，表示这是最简单的无穷集。我们熟悉的那些运算法则对于这个集合是无效的，这与我们对无穷大的理解是一致的，伽利略那部著作中的辛普利西奥也发现了这个问题。例如，从集合的本质我们可以得到下面这些运算法则：

$$\aleph_0 + 1 = \aleph_0$$

$$\aleph_0 + \aleph_0 = \aleph_0$$

$$\aleph_0 \times \aleph_0 = \aleph_0$$

利用集合论，我们可以毫不费力地解决让伽利略头疼不已的那个平方数与正整数的难题。我们知道，可以根据两个集合中的元素是否可以构成一一对应的关系来判断它们是否等势。在解决平方数和正整数的这个难题时，我们同样可以利用这个方法，将它们一一配对，即每个平方数对应一个正整数。由于我们可以完成这个步骤，说明这两个集合是等势的，都是$\aleph_0$。因此，无穷集与自身的一个子集等势的奇怪现象就可以解释了。我们在指南针指针方向与季节的例子中已经发现，无须知道集合有多少个元素，也可以确定它们是否等势。

如果我们根据自己在周围世界中获得的体验来理解数学过程，就会导致这样的问题。我们难以理解无穷集的性质，是因为我们以为它们具有与有穷数字（尤其是现实世界中数量有限的物体）相类似的特点。然而，尽管集合可以帮助我们理解数字的含义，但是集合不是数字，而是一种数学建构。只有清楚地理解集合是一种截然不同的实体（它与数字的关系可以帮助我们理解数字），我们才能正确理解无穷集的奇怪特性。

在集合论的帮助下，康托尔成功地定义了自然数的无穷性，即$\aleph_0$。我们也许会认为，集合论对无穷大的研究已经到了极致。但是，作为数学家，康托尔绝不愿意不加深究就轻信任何结论，这也是他在“$\aleph_0$”后面附上一个0的初衷。这只是最简单的无穷，但是他还没有证明，包含整数在内的所有数字构成的集合，是否也与之等势。因此，康托尔决定继续研究下去。这一次，他使用的是非常简单易懂的数学证明。现代数学证明往往包含一页又一页的方程式。20世纪，安德鲁·怀尔斯证明著名的费马大定理的过程超过100页纸的篇幅。然而，不用任何方程，我们也可以基本理解康托尔的无穷集合证明。

证明过程必须注重严谨性，在用数学语言描述时，仅仅使用几张图表肯定是不够的。在介绍康托尔的证明时，我将对证明过程略做压

缩处理，但是它们仍然非常直观，古希腊几何学家肯定也会欣然接受。不幸的是，与康托尔同时代的人却持有不同的看法。

康托尔考虑的第一类数字是有理分数。他想象把所有正有理分数填入一张表中，使分子从左到右逐项加1，分母自上而下也逐项加1。于是，他得到下面这张表：

1/1	2/1	3/1	4/1	5/1	6/1	7/1	8/1	9/1	10/1	…
1/2	2/2	3/2	4/2	5/2	6/2	7/2	8/2	9/2	10/2	…
1/3	2/3	3/3	4/3	5/3	6/3	7/3	8/3	9/3	10/3	…
1/4	2/4	3/4	4/4	5/4	6/4	7/4	8/4	9/4	10/4	…
1/5	2/5	3/5	4/5	5/5	6/5	7/5	8/5	9/5	10/5	…
1/6	2/6	3/6	4/6	5/6	6/6	7/6	8/6	9/6	10/6	…
1/7	2/7	3/7	4/7	5/7	6/7	7/7	8/7	9/7	10/7	…
1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	9/8	10/8	…
1/9	2/9	3/9	4/9	5/9	6/9	7/9	8/9	9/9	10/9	…
1/10	2/10	3/10	4/10	5/10	6/10	7/10	8/10	9/10	10/10	…
…	…	…	…	…	…	…	…	…	…	…

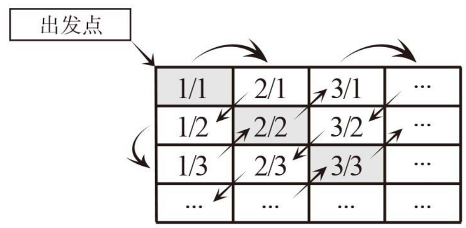
显然，这张表是无法完整画出来的，因为它是一张无穷大的表。但是，我们可以看出它的规律。表中包含所有的有理分数，并且数字1将沿着对角线方向出现无数次。接下来，康托尔需要在表中找出一条可以到达所有位置的简单重复路径，才能证明表中的有理分数集与自然数集等势。也就是说，他需要制定一些法则，即算法，来帮助他通过一个简单的程序覆盖表中的所有方格，比如：

1. 从左上角出发。
2. 向右前进一步。
3. 向左下方运动至表的边缘。
4. 向下前进一步。

5. 向右上方运动至表的边缘。

6. 重复第2步及后续步骤。

通过这个程序，最终可以走过表中所有有理分数，并且不会有任何遗漏。



可以采取的路线不止一种，但重要的是，沿着康托尔建立的这条路线，我们可以一步一步地走遍所有方格。我们每迈出一步，就会走过一个方格。我们需要做的就是将这些步骤与整数配对，从而按部就班地建立起一一对应的关系。也就是说，总体来看，表中的有理分数集与整数集等势。因此，有理分数集的元素个数是$\aleph_0$。

这个结论自然不会让我们大吃一惊。毕竟无穷大非常特殊，而且我们知道下面这个等式是成立的：

$$\aleph_0 \times \aleph_0 = \aleph_0$$

直觉告诉我们，有理分数集符合这个规律是有道理的。但是，这并不意味着所有数学现象都是合理的。当康托尔使用同样的方法研究另一个数集时，他无比震惊地发现结果竟然大相径庭。

想一想，0—1之间有哪些数字。（康托尔研究的其实不是这些数字，但是0—1的数字考虑起来最简单。）这里说的“数字”指什么呢？不仅仅是整数（如果我们说的0—1这个范围包含边界，那么共有两个整数），也不仅仅是有理分数（0—1之间的有理分数就是第195页表格第一列中的所有数字，也就是分子是1、分母是各个整数的所有分数。它们是所有分数的一个势为$\aleph_0$的子集）。除了这些数以外，还有无理数，即与2的平方根相类似的数，但我们在这里讨论的无理数数值都在0—1之间。

从本质上讲，康托尔考虑的其实就是0—1之间的所有小数（即“实数”），而且包含这个范围内所有可能的数字。要使用上面那个方法，我们必须将表格打乱重排，否则小数的开头就会有无数个0，无论多大的纸也写不下这些0。重新排列之后，我们可能会得到下面这张表格：

0.**6**47 204 217 589 579 326 910 461 74…
0.3**1**5 729 571 037 472 748 274 373 33…
0.61**9** 572 749 457 117 572 995 274 21…
0.911 **3**67 284 573 823 565 633 612 21…
0.162 2**3**4 627 512 255 579 823 671 49…
0.236 25**0** 558 519 367 215 249 473 25…
…

下面这个步骤充分展示了康托尔的天才。他按照每次后移一位的方式，从各个数中选出一个数位加粗。然后，他把这些加粗的数字排列起来，再逐项加上1（如果原来的数字是9，加1之后就会变成0）。

这样，这些数字串就可以构成一个0—1之间的小数。例如，我们可以从上表得到下面这个小数：

0.720 441 784 983…

这个数字非常有意思。它与康托尔表格中的第一个数不同，因为它们的第一个小数位不同；它与表格中的第二个数不同，因为它们的第二个小数位不同；它与表格中的第三个数不同，原因同上。以此类推，它与表格中的所有小数都不相同。也就是说，我们得到的这个小数并不包含在上表中。

如果我们可以成功地写出0—1之间的所有数字，我们就可以把这些数字与正整数逐个配对，从而证明小数集与自然数集等势。但事实上，我们无法写出所有小数。康托尔告诉我们（并给出了严格证明），0—1之间的数字比整数多。这个集合的势更大，是更大的无穷大。

接下来，康托尔把探索的触角伸向其他维度。他把这个更大的势称作$\aleph_c$，因为它是0—1之间的连续统的势，也就是数轴上0—1之间所有点构成的集合的大小。然而，我们经常描述的是二维平面或者三维空间里的点，而数学家通过假设，可以轻松自如地考虑任意维度。这些无穷大是否适用于这些假设的维度呢？

我们同样可以在几乎不使用数学工具的条件下，轻松地把康托尔接下来的证明过程解释清楚。我们通常会使用一组坐标（也就是我们前面讨论过的笛卡儿坐标系）来定义二维平面上的点，这些坐标可能是坐标图上的x和y，也可能是地图上的经度和纬度。因此，边长为1的正方形区域中的所有点都可以用两个0—1之间的实数来定位。

直觉告诉我们，既然$\aleph_0 \times \aleph_0 = \aleph_0$，我们似乎就可以将0—1之间的连续统的势按比例增大，从而将同样的法则应用于正方形中的所有

点。但是，这并不是一个有效的证明。康托尔发现，把表示某个点的坐标的两个数放到一起，使它们的各个数位交错排列，就可以得到一个独一无二的小数，而且可以用这个小数确定这个点。这样一来，正方形区域内所有点的势就再次变成了0—1之间所有小数的势。通过交错排列更多的数位，我们可以延展至任意多的维度。就这样，连续统的无穷性再次体现在正方形中的所有点上。

考虑数轴上0—1之间数字的无穷性，可以得出一个在研究无穷大时经常会让我们感到头晕眼花的悖论。根据康托尔的证明，我们知道有理分数集与整数集等势。接下来，我们让另外一个分数集，即1/2，1/4，1/8，1/16…这个数列，与有理分数并列，很容易证明它们也等势。此外，我们已经知道这个无穷级数的和是1。做好这些准备工作之后，好戏就要开场了。

假设我们给数轴上的每个正有理分数发一把伞，以防止它们被雨水淋湿。这些伞都是简单的T形。第一把伞的T形结构在数轴上占据1/2个单位，第二把伞占据1/4个单位，以此类推。一旦所有的正有理分数都撑起伞，整个数轴就会全部被遮盖在雨伞之下。伞的T形结构向两边伸出的幅度相同，也就是说，第一把伞将遮挡住左右各1/4个单位中的所有数字。请注意，由于伞遮挡的都是有理分数，将第一把伞所在的点（同样是一个有理分数）加减1/4都会到达另一个有理分数。

到目前为止，你没有发现任何问题吧？每把伞都立在一个有理分数所在的点上，同时向两边展开，伸到其他有理分数所在的位置。别忘了，我们给每个正有理分数都发了一把伞，因此伞与伞之间至少会相互接触，大多数情况下还会发生重叠，从而把整个数轴都遮挡起来。

也就是说，我们利用雨伞把数轴上0至无穷大的部分全部遮盖起来。现在，再想一想伞的宽度，它们的宽度构成了无穷级数1/2 + 1/4 + 1/8 + …。在不发生重叠的情况下，这些雨伞最多可以覆盖数轴的1个

单位，在发生重叠时，覆盖的长度就更小了。总宽度仅为1的物体集合竟然把无限延伸的直线都遮盖住了，你是不是感到困惑不解啊？这就是无穷大给我们的大脑带来的冲击。

然而，尽管康托尔取得了这些成就，他却始终无法证明一个发现的正确性。他最终精神崩溃，或许就是出于这个原因。康托尔认为，无穷应该是分等级的，整数的无穷等级最低，其次是连续统的无穷等级，即$\aleph_c$。这个观念在他的心目中几乎达到了宗教的高度，事实上他把终极无穷与上帝联系到了一起。但是，他没有办法证明连续统的无穷等级是$\aleph_1$。在$\aleph_0$和$\aleph_1$之间可能还存在其他无穷等级。直到去世，康托尔也没有发现这是一项毫无意义的研究。后来，有人利用数学方法证明，这个被称作连续统假设的断言的正确性根本无法确定。

这个结论是由数学家库尔特·哥德尔通过他的不完全性定理证明的。不完全性定理是所有数学家的噩梦。该定理认为，任何一个形式系统，只要包含了一阶谓词逻辑与初等数论，就必然存在一个命题，它在这个系统内既无法被证明为真，也无法被证明为伪。有时，即使在同一个世界内，我们可以应用的法则也可能不止一套。前面讨论的平面和曲面上的平行线的特性就是一个例子。但是，当我们试图使用数学工具探索现实世界时，就必须使用一套固定的法则。

哥德尔的研究实质上是要证明任何系统中都会存在某些结果无法确定的问题。也就是说，从根本上看，数学是不完善的。如果把哥德尔不完全性定理简化，就与我们在前文讨论的罗素悖论非常相似了。罗素悖论给出的命题在应用数学系统的法则时会产生无法解决的问题。哥德尔成功地证明康托尔的连续统假设与集合论公理系统不矛盾——这个假设有可能是正确的，但是他无法证明。后来，另一位名叫保罗·科恩的数学家，证明了集合论与连续统假设彼此独立。换句话说，即使该假设不是真的，对集合论也不会产生任何影响。

从本质上讲，他们的研究表明，只要现行的集合论公理系统不做修改，就不可能证明连续统的无穷级别就是高于$\aleph_0$的$\aleph_1$。哥德尔本人也着重强调了这一点，他说连续统假设的真实性肯定无法确定。他还说：“根据现在已知的集合论公理系统无法确定它的真实性，只能说明这些公理无法完整地描述现实世界。”

我们知道，公理是为某个数学分支奠定基础的基本假设。数学家必须假定这些“已知”条件是真实的，然后以它们为基础，开始搭建数学结构。运用数学证明时一定会使用这些公理，但是这些公理毕竟是假设，假设肯定会引起人们的质疑，因此这些公理到底正确与否，难免引发争议。

集合论是康托尔所有无穷大研究（更不用说这位数学家眼中的数字基本概念及运算法则）的基础，集合论自身的基础则是ZFC公理系统。其中Z和F分别指将康托尔的集合论研究成果整理成形的数学家策梅洛（Zermelo）和弗伦克尔（Fraenkel），C代表选择公理（把这条公理单独列出的原因马上就会揭晓）。系统中的8条公理对于20世纪的数学界而言非常熟悉：

1.存在性公理。至少存在一个集合。基数源自空集，即没有任何元素的集合。但是，首先必须有集合存在。

2.外延公理。当且仅当两个集合有同样的元素时，这两个集合相等。这条公理具有数学公理的典型特点：表面上是一个显而易见的命题，但是，要让数学有据可循，它又不可或缺。

3.分类公理。对于所有集合与所有条件，都有一个集合与之对应，且该集合的元素正好是原集合中符合该条件的所有元素。换言之，从一个集合中选择一些元素，无论如何选取，所选择的元素都可以构成一个集合。例如，在所有大于1的自然数这个集合中，利用“除自身和1以外没有因数”这个条件，就可以得到另外一个集合，即素数集。

4.无序对公理。对于任意两个集合，都存在第三个集合将前两个集合包含其中。也就是说，可以由两个集合得到第三个集合。

5.并集公理。已知多个集合，则存在某个集合包含属于已知多个集合之中至少一个集合的所有元素。

6.幂集公理。对于任意已知集合，都存在一系列集合，包含已知集合的所有子集。

7.无穷公理。存在一个集合，包含空集和所有非空子集。

8.选择公理。对于任意集合，我们都有办法从该集合的所有非空子集中选择一个元素。

这些公理大多比较可靠，而且不会导致麻烦。但是如果我们处理的是无穷集，最后那条公理，也就是第8条公理，就会成为ZFC系统的大麻烦。隐患就在于“办法”这个词。当然，对于一个已知集合，我们肯定可以从中随机选择一个元素，但是“随机选择”并不能被视为一个有效的数学方法。我们可以不考虑任何特殊原因，从一系列物理对象中随机选择一个，但是利用数学方法完成随机选择的难度非常大，因为很难定义到底如何选择才算真正做到随机，除非集合的元素数量已知。

对于有穷集，我们甚至无须做到随机选择，比如，我们可以采取“选择集合的第一个元素”的方法。但是，整数是所有数字的一个子集，它们沿着正负方向无限延伸。如果我们需要从这个子集中选取一个数字，应该如何做呢？也许“选择中值”这条法则可以帮助我们完成任务，但是无穷集的中值真的那么显而易见吗?

好消息是，我们现在有办法“修复”ZFC系统，为连续统假设这个问题给出一个确切的答案。大多数数学家都认为最好的办法是“力迫法”或“内模型法”（这个方法还有一个上口的别称——V = ultimate L，

即集合论宇宙终极可构成集类）。唯一的问题是，力迫法告诉我们连续统假设是错误的，而内模型法则认为连续统假设是正确的。

这两个可能的结果充分显示了数学的本质和它与现实世界的关系。集合论绝对是我们每天都要使用的实用数学的基础之一，然而集合论自身在选用公理这个方面却具有随机性。一条道路通向光怪陆离、令人向往的数学世界，另一条道路则更贴近我们心目中的现实世界。就纯粹数学而言，这不是问题，只能说明我们使用的是两个不同的数学系统，就好比有的数学系统是建立在维度多达数千而且与现实世界毫不相干的基础之上。但是，作为科学的基础，我们还是希望找到一条具有唯一性和确定性的数学道路。

尽管在科学研究中应用数学工具，会因为集合论自身的问题而受到严重影响，但是康托尔的无穷性研究并不会给我们带来明显的麻烦。我们知道，如果方程涉及变化，而且某些值可以通过合并无穷多个无穷小的部分的方式推导得出，牛顿、莱布尼茨及其后来者在进行微积分运算时就会使用潜无穷的概念。在这种情况下，无穷的概念（至少是亚里士多德提出的潜无穷概念）具有让人无法抗拒的价值。我们更不清楚的是，如果脱离了彻底抽象化的数学世界，康托尔的实无穷是否还有任何实际意义。

答案非常简单，到现在为止，我们还不知道无穷在现实世界中到底有没有意义。宇宙可能是无穷无尽的，大爆炸理论并不能彻底否定这种可能性。我们只知道可观测宇宙的直径大约是900亿光年，这个数字是对光自宇宙诞生以来传播的距离与相同时间里宇宙膨胀的速度加以综合之后得出的结论。但是，无论我们这个宇宙是独一无二的，还是无穷无尽的宇宙海洋中不起眼的一滴水（很多现代大爆炸模型认为，在更大的多元宇宙中发生过多次大爆炸），宇宙的膨胀都有可能是没有限度的。

从某些方面看，无限宇宙似乎比有限宇宙更有吸引力，因为有限宇宙会让人们情不自禁地遐想：宇宙边界之外是什么？但是，数学家已经找到了一个可能的答案：即使是有限宇宙，也可能完全没有边界。

这个答案似乎与我们的直觉不符，因为这在常见的三维空间中是很难想象的，但是我们可以轻松地找到一个二维类比对象——月球表面。（我选择月球表面而没有选择地球表面，是因为地球表面上的海洋会破坏它的连续性。）月球上有一个有限空间，即它的表面，而且这个有限空间没有边界。我们可以朝着任意方向一直走下去，也不会走到月球的边缘。站在月球上看，月球表面似乎是一个平面。把平面折向第三个维度，让边缘结合到一起，平面的边界就消失了。要在宇宙中想象出类似效果，我们需要把宇宙的体积折向第四个维度。在折叠之前，我们在某个位置上跨出一步就相当于从宇宙的一个侧面跨出去，但是在折叠之后，跨出去的这一步会让我们从宇宙的这一面进入与之相对的另一面。

现实世界中另一个可能真正具有无穷性的事物是时间。亚里士多德认为时间没有尽头，可以被视为一个无限膨胀的过程。有的宇宙学理论认为宇宙也没有起点，尽管更受欢迎的大爆炸理论揭示了宇宙的起源。但是，有人认为，如果宇宙膨胀到一定程度，之后再也不会发生任何变化，宇宙就走到了尽头。在这种情况下，我们有理由相信时间已经不存在了，因为时间的流逝将没有办法表示。

亚里士多德还相信时间和空间都可以被分割成许多无穷小的部分（亚里士多德不是原子论者）。现代物理学家大多认为这个观点可能不正确。仅仅因为物理现实的其他方面可以量子化（可以分割成离散粒子），时间和空间就一定也如此吗？考虑到万有引力已经被引入量子框架，他们更加确信这个观点不正确了。表示空间粒度效果最好的

备选量度就是普朗克长度和普朗克时间，这两个单位的值是由光速c、引力常量G和普朗克常量h三个基本常数决定的。

当人们建立这两个普朗克单位时，它们被视为宇宙的基本属性，而不是人为创造的两个概念。普朗克长度等于$\sqrt{hG/c^3}$，大约是1.6×10^{-35}米，普朗克时间是$\sqrt{hG/c^5}$，大约等于5.4×10^{-44}秒。还有一个普朗克单位是普朗克质量，即$\sqrt{hc/G}$，大约等于2.2×10^{-8}千克。虽然人们很少提及普朗克质量，但这是唯一一个与我们体验过的事物具有可比性的普朗克单位。（这些单位的值非常小，人们从来没有用过其他任何方式来衡量这些值。）如果时间和空间可以量子化，那么我们有可能（尽管无法确定）使用这些值来测算它们的粒度。（普朗克质量肯定不是质量的最小单位，它大约是电子质量的10^{22}倍。有人认为，普朗克质量可能是基本粒子的最大可能质量。）

有趣的是，可能出现真正的无穷的一个领域是量子物理。原子、电子等量子的属性不同于我们在“宏观”世界中熟悉的各种事物的属性。例如，量子的一个特性叫作自旋（量子自旋并不是指真的自旋，而是借用自旋物体来比喻某种属性），但是我们无法用具体的值来表示它。在任一特定方向上测量，我们会发现量子自旋的方向可能“向上”，也可能“向下”。

在测量之前，我们无法预知测量结果，因为粒子的自旋没有实值。其实，自旋就是一种叠加状态，比如，粒子有27%的可能性向上自旋，有73%的可能性向下自旋。也就是说，如果我们在某个特定方向上重复测量，测量结果是“向上”和“向下”的概率分别是27%和73%。但是，我们只能测量概率（这要归功于薛定谔方程），而无法预知测量结果。

在叠加状态下，自旋可以被视为一种方向。我们可以把它想象成箭头，箭头朝上表示上旋的可能性是100%，箭头朝下则表示下旋的可能性是100%。在叠加状态下，箭头指向某个中间方向。（可能性为各占一半，箭头指向水平方向。）由此可见，我们实际上得到了一个实数——一个无穷小的数，其表现形式是叠加状态的方向。因此，以这些粒子自旋为基础的量子计算机的潜能远远超过二进制的传统计算机。如果$\aleph_0$准备入侵现实世界，也许它会通过量子位，甚至只能通过一种间接的方式来实现这个目的，因为我们永远无法知道它的确切值。但是，这种入侵是真实的，因为它会对测量结果产生直接影响。

然而，有一件事肯定是真实无误的。尽管对于数学家而言，无穷就是一个雅致的玩具，但是在科学研究中却经常会招致难以解决的麻烦。在物理学领域，例如在研究前面讨论的电子反冲时，无穷大概念经常会发挥更大的价值。无论我们研究的是黑洞内部结构，还是反冲电子这种常见现象，无穷都有可能悄无声息地出现在我们眼前。电子还会导致另外一个问题，因为人们认为电子是没有维度的点粒子。但是，这就意味着随着我们越来越接近电子，电场强度就会趋近于无穷大，而离电子最近的肯定是它自己。

当用量子电动力学解释光与物质粒子之间发生的电磁相互作用时（这种电磁相互作用对于我们所有的日常体验来说几乎都是不可或缺的），无穷引起了无数麻烦，电子与自身电场发生相互作用产生的自能就是其中之一。然而，人们普遍认为，就预测观测值而言，量子电动力学是迄今最成功的理论。那么，它是如何处理无穷问题的呢？答案是重正化，归根结底就是用观测值取代那些荒谬的无穷值。

在遇到无穷时，物理学家并不总是束手无策，因为他们随时可以借助微积分中的潜无穷概念来化解危机。但令人尴尬的是，物理学理论似乎随时会抛出足以导致麻烦的实无穷概念。物理学家马克斯·泰格

马克认为，如果我们支持那些容许实无穷概念存在的理论，就等于给未来的物理学制造麻烦。

他特别指出，宇宙膨胀说就是这种理论的一个代表。宇宙膨胀说是帮助早期大爆炸理论解决某些问题的“补丁”。该学说认为，大爆炸发生之后，宇宙空间以远超光速的速度急剧膨胀，我们现在可观测到的宇宙从此开始了它的生命历程。现行的宇宙膨胀说与观测结果的一致程度比较高。（从某种意义上讲，这也是理所应当的，因为宇宙膨胀说的创立初衷就是实用，为了与新数据吻合，又修改了若干次。但是，自现代宇宙膨胀说建立之后，已经有很多观测结果迟迟不能得到合理的解释。）

泰格马克称，麻烦的根源就在于膨胀说秉持宇宙体积可以无限膨胀的观点。根据这个观点，最终将会形成一个空间无穷集，将所有可能的物理情况都包含其中，导致膨胀说在诸多领域里完全丧失做出明智预测的能力。如果一切都有可能，我们就无法准确预测任何结果，科学的意义也会遭到严重破坏。从这个角度看，宇宙膨胀说与致命计算机病毒有几分相似，如果听之任之，所有的科学理论都将遭遇灭顶之灾。

泰格马克指出，就像橡皮筋因为原子数量有限而无法无限拉伸一样，根据时空的量子性质，宇宙的膨胀也应该是有限度的；而且，如果物质真的具有连续性，那么这个说法基本上就是对的。他认为，有了这个限度，一切问题都将迎刃而解。无论是密度无限大的黑洞奇点，还是阻碍量子引力理论发展的数学难题，都不再是问题。他大声疾呼：我们不需要实无穷！

最后，泰格马克说：“我们物理学家面临的挑战是找到这个简便有效的方法，用不包含无穷的方程描述真正的物理定律。我们必须先对无穷提出质疑，才可能积极投身这项探索活动。我认为，我们有必要

把它赶出物理学界。”尽管泰格马克的话有些特立独行，但是他的思想可能代表了科学的一个新起点。

与令人尴尬的物理学领域的无穷不同，康托尔研究的数学领域的无穷对日常生活与科学研究从未产生任何重大的影响。从研究数字与现实之间关系的角度看，我们更想知道哥德尔的研究成果，以及选择公理因为自身问题而导致的随机性到底会产生什么样的影响。集合论是数学的基础，但它自身却有一个有趣的缺陷。或许这些新发展的最大意义是告诉我们，将数学视为现实世界的直接基础，会带来一定的风险。果真如此的话，就意味着现实也具有随机性。

尽管直到康托尔去世，他的无穷理论也没有得到广泛应用，但是物理学却开启了一个新的发展方向，使数学的核心地位得到了史无前例的巩固。从此以后，人类对世间万物的认知，以及人类的日常生活，几乎都将因此而发生改变。

第13章
爱因斯坦：量子物理与抽象数学

Are Numbers Real?

学校里有一个非常奇怪的现象：在中学毕业之前，学生们学习的物理学知识与19世纪末的物理教学内容完全相同。对于1900年以后的物理学，他们可能略知一二，但是，除非他们上大学之后继续学习物理，否则就不大可能了解这门科学在20世纪到底取得了哪些发展（更不用说21世纪的物理学了）。

其他学科则几乎不可能发生这样的情况。例如，如果英语课只介绍19世纪90年代之前的文学作品，大家肯定会觉得十分奇怪。但是，从1900年至今，文学领域的变化远比不上物理学，因为20世纪产生了两大革命性理论——相对论和量子论，1900年之前的所有物理学知识都无法逃脱被修改或者被摒弃的命运。作为新科学的组成部分，这两大理论都从数学那里获得了重要的推动力。

相对性这个基础概念要追溯至伽利略时代，痴迷于研究对称的人（参见下一章）肯定认为相对性与对称的本质——不变性有关。所谓不变性，是指不会随周围环境的变化而变化的力、物体或物体的某种特性。也就是说，即使发生某个特定变化，某事物的活动方式与结果

仍然保持不变。在伽利略提出的相对性原理中，不变与稳态运动有关。

众所周知，伽利略曾经对人们深信不疑的几个古代信条提出过严厉批评，他提出相对性原理也是出于这个目的。据传，伽利略在一次外出游湖时证明了相对性这个概念。当时，他和几位朋友在皮耶迪卢科湖上泛舟，船以某个固定速度前进。伽利略问谁身上有比较重的物体，他的朋友斯泰卢蒂拿出自己的钥匙，递给伽利略。这把钥匙是铁制的，很大很结实，如果丢失，将很难配制。伽利略接过钥匙，朝着正上方用力抛出去。因为6名船工正在奋力划桨，因此船前进的速度非常快。斯泰卢蒂赶紧向船后方跑去，希望能接住掉下来的钥匙。但是，另外几名朋友拉住了他。最终，钥匙垂直掉下来，落在伽利略的大腿上。

伽利略通过这个“实验”证明了他提出的观点：在稳态运动的物体上，例如那条船，我们可以顺利完成任何物理实验，实验过程中不会受到船以外世界的任何影响，实验结果与我们在静止状态下完成的实验没有任何不同。如果伽利略坐在岸上不动，斯泰卢蒂知道钥匙肯定会垂直上升然后垂直下落。但伽利略却知道，在船上抛钥匙也会得到同样的结果。所谓“相对性”，就是伽利略发现相对钥匙而言，船并没有移动。只有相对于其他事物，才可以发现或探查到稳态运动。所以，伽利略规定了“船以外世界”这个条件。比如，与湖岸相比，船正在运动，并且我们可以通过实验方便地探查出船的运动状态。

伽利略的研究并没有使用复杂的数学知识，但即便如此，学校的科学课程也往往不会把伽利略的相对性理论纳入教学内容。似乎只要提起“相对性”这个词，就会让那些教育家们感到害怕。我们会把结果教给孩子们，但是我们不会认真解释相对性这个重要概念的真正含义。我知道，大多数教师都会辩解说，他们之所以没有深入讨论，是因为这些内容太难了。当然，如果讨论广义相对论和量子理论，对知

识水平的要求肯定不是中学生可以达到的。但是，如果只要求他们掌握相对性和量子理论的概念，就不会有太大的难度了。事实上，我到学校做报告时，发现学生们理解这些概念的速度比成年人更快，这可能是因为年轻人更习惯于接受新奇的东西。

然而，爱因斯坦改进伽利略相对性原理的第一个成果——狭义相对论，对数学水平的要求并不是太高，所以实在没有理由将它排除在中学课程之外。的确，理解狭义相对论需要掌握数学知识，而且这些数学知识与牛顿定律中使用的我们习以为常的数学知识有所不同，但是，两者之间的跨度是可以接受的。此外，狭义相对论是一门有极强吸引力的科学。比如，一讲到时空穿梭机，学生们个个都正襟危坐，认真听讲。因此，中学不教授狭义相对论是没有任何道理可言的。

从本质上讲，爱因斯坦的研究实际上是对伽利略的相对性概念进行了拓展，并将麦克斯韦揭示的光的本质也纳入其中。麦克斯韦的研究表明，速度是电磁波的一个决定性特征。根据伽利略的相对性原理，如果我们与某个正在运动的事物并肩前进，而且速度相同，对我们来说，这个“正在运动的事物”就是静止不动的。因此，在伽利略把斯泰卢蒂的钥匙朝正上方扔出去后，钥匙会掉落在他的大腿上。但是，如果这个原理同样适用于光，那么与光并肩前进就会改变光的速度。一旦光的传播速度不等于光速的定义值，光将不复存在。

一般人可能会认为麦克斯韦错了，但是爱因斯坦十分赞赏麦克斯韦的证明过程。他由此推断光与其他事物不同，即使我们与光做相对运动，光的速度也不会发生改变。这似乎是一个无关紧要的变化，但是一旦把这个变化放入研究运动的基础数学（始于伽利略和牛顿）中，就会对现实的本质产生令人惊讶的影响。

所谓的“光钟”就是一个非常简单的例子，可以帮助我们了解这种变化的影响力。光钟这种装置没有钟表的“滴答”声，取而代之的是一束光在两个水平镜面之间不断反弹。当上下传播时，光束始终在一条

直线上。假设把光钟放到一艘透明的宇宙飞船中，然后我们在地球上用一台超级望远镜观察它。如果飞船没有与地球做相对运动，那么地球上的人与飞船上的人从光钟上看到的时间应该是一模一样的。在这两个人看来，光钟都没有移动，因此光束在镜面之间反射时肯定与镜面垂直。

现在，我们假设飞船正在以一个很快的恒定速度相对于地球运动。伽利略的相对性原理告诉我们，对飞船上的人而言，飞船里没有发生任何变化，光钟在他们眼中没有运动，那束光继续沿着直线稳定地上下传播。但是，地球上的人却看到了某些变化。现在，假设那束光刚好从光钟顶部的镜面上反射出去，在它到达底部镜面的这段时间里，飞船将发生侧移。于是，光将不再垂直向下传播，而是沿着一条更长的斜线传播。在光返回光钟顶部的镜面的过程中，同样的现象将再次发生。也就是说，光将成“之”字形传播。

如果伽利略的相对性原理是正确的，情况就会大不相同。假设我们坐在皮耶迪卢科湖岸边，观看湖面上伽利略乘坐的那条船。船上有一个与光钟相似的东西，正在由上向下发射一串钢珠。在船上的人看来，“光钟”没有动，钢珠将沿着垂直方向向下运动。但是，我们坐在岸上会看到船与钢珠都在发生侧移，因此我们会将钢珠的两种运动加到一起，计算出一个新的速度。但是，如果爱因斯坦是对的，也就是说，无论我们以何种方式与光做相对运动，光的速度都会保持不变，那么地球上的人在观看光钟时，看到的情景就会与飞船上的人不同。光传播的距离增加了，但是速度不会改变。

这就意味着肯定有其他因素发生了改变，光才可以按时抵达目的地。爱因斯坦通过数学计算，发现必然会产生三种效应。从地球上看，飞船上的时间变慢了，运动距离缩短了，飞船的质量增加了。根据狭义相对论，空间和时间再也不被视为两个完全独立的实体，研究两者的结合体——时空，变得非常有必要。

事实上，要解决这个问题，只需要学习一点儿古希腊几何学和计算平方根的数学知识（所以，中学生应该学习狭义相对论）即可。爱因斯坦根据光钟的几何结构，再通过几个思想实验，计算出通常被称作伽马（γ）的关键因子就等于$\sqrt{(1-v^2/c^2)}$，其中v是观察者观察到的运动物体的速度（本例中就是宇宙飞船的速度），c指光速。

我们还是以光钟为例。如果飞船上的时间变化量为t，那么对于地球上的观察者而言，这段时长是t/γ，即$t/\sqrt{(1-v^2/c^2)}$。这个表达式看似涉及高深的数学知识，但其实非常简单。如果v等于0，即飞船保持静止，则v^2/c^2等于0，除数为1，最后得数就是t。此时，对于飞船与地球上的人而言，时间流逝的速度是相同的。

但是，随着v越来越接近光速（c），t被一个越来越小的数除，因此从地球上观察，这段时间就会越来越长。如果γ等于1/2，则这段时间等于$2t$。如果γ等于1/4，则这段时间是$4t$。对于飞船上的人而言，时间流逝的速度没有任何变化，但是对地球上的人来说，飞船上时间流逝的速度变慢了，当飞船上流逝的时间为t时，地球上的时间已经流逝$4t$了。

最让人感到奇怪的是，相对性是完全对称的。地球上的人往往把地球表面定义为“静止”，但是这个判断具有主观随意性（尽管通常比较方便）。毕竟，地球不仅绕着轴线自转，沿着轨道围绕太阳公转，还与银河系一起在太空中高速遨游。但是，我们往往不会这样想，而是采用了伽利略站在游船上的视角，认为船是静止的，而周围世界正在向我们身后运动。与之类似，从飞船乘客的视角来观察，飞船根本没有动，而地球正在向远方高速飞去。至于选择其中哪方作为固定点（科学家常常称之为“参照系”），我们没有一定之规。因此，如果飞船上的乘客可以看到地球上的光钟，他们就会发现因为地球正在远离飞船，所以地球上的时间变慢了。

我在这里并不是要讨论以接近光速飞行会产生哪些结果，但是我必须简要说明在著名的“时间膨胀”效应的例子中为什么没有这种对称性。在这个思想实验中，两个双胞胎姐妹参与了一项飞船任务。其中一个人是任务主管，另一个人是宇航员，后者要乘飞船以接近光速的速度飞行，执行一个长期任务。假设任务开始时两姐妹正好30岁，至任务结束时，“宇航员”感觉时间过去了5年，但是她发现留在地球上的姐姐却长了10岁。

如果相对性像上文所说的那样具有完全对称性，那么这个思想实验似乎不应该出现这样的结果。在飞船以恒定速度飞离地球时，对称性应该始终发挥作用。但是，这种对称性随后发生了改变。在某个时刻，宇航员必须向飞船施加一个作用力，让它减速，然后加速返回地球。抵达地球时，她需要再施加一个作用力，使飞船的运动速度等于地球的运动速度。这个变化只发生在飞船上，而没有发生在地球上。在它的影响下，时钟被校准了，所以宇航员感觉时间只过去了5年。

毫无疑问，狭义相对论的数学运算经常给出一些令人惊讶的结果，包括时间膨胀、没有质量却有动量的粒子（例如光子），以及把能量与质量联系到一起的终极方程$E = mc^2$。然而，一般而言，其中用到的基础数学知识对于高中生来说并不是很难。但是，广义相对论的情况有所不同。在研究广义相对论时，即使爱因斯坦也要在数学上向人求助。尽管关键的几个方程看似非常简单，但是其中隐藏了极大的复杂性，只有在特例中才可以求解，而不是在任何情况下都能找到答案。

广义相对论将加速度和万有引力纳入其中（因此，与狭义相对论相比，广义相对论中的相对性具有普遍性），但是它所涉及的基础知识并不是特别难。广义相对论始于现在被称作等效原理的想法（爱因斯坦称之为“最快乐的思想”）。

爱因斯坦产生这个想法的时候还是一名业余科学家，在瑞士专利局上班。（爱因斯坦在这里完成了狭义相对论的全部工作，但是他在完成广义相对论的主要研究前，一直未获得学术界的认可。）他后来说："当时，我正坐在伯尔尼专利局的椅子上，突然我的脑海里闪过一个念头：如果一个人做自由落体运动，他肯定不会感受到自己的体重。我吓了一跳。这个简单的想法让我久久不能忘怀，在它的驱动下，我开始研究万有引力理论。"

不熟悉广义相对论的人，无法从爱因斯坦的这番话中看出他当时到底想到了什么。爱因斯坦其实是在举例子。如果有人从非常高的建筑物上掉下来，他就会加速下落，而且加速度是一个标准值，由地球与他的身体之间的万有引力决定。但是，与此同时，他会有失重的感觉。就像国际空间站里的宇航员一样，他会感觉自己飘浮在空中。当然，你在从高处掉落时会使周围空气发生振动。因此，想象你在箱子里，与箱子一起做自由落体运动，效果可能会更好。实际上，国际空间站的宇航员之所以有失重感，唯一的原因是他们正在下落。

值得注意的是，空间站高度上的地心引力强度与地面其实非常接近，大约是正常地心引力的9/10。造成宇航员失重感的原因是空间站正在垂直下落，就像从建筑物上掉下来的人一样，正在做自由落体运动。两者之间唯一的区别是空间站还在向侧方高速运动，因此，尽管它一直在下落，却怎么也到不了地面。这是进入运转轨道之后必然产生的结果：一直在下落，却怎么也到不了地面。

因此，那些飘浮在空间站中的宇航员就是一个平常而真实的例子，他们向我们演示了爱因斯坦的想法：做自由落体运动的人感受不到自己的体重。在他们加速下落的时候，他们本来可以感受到的引力被抵消了。相比之下，如果我们下落的加速度超过了引力的作用，我们就会产生引力场正在将我们朝相反方向拉扯的感觉。大家想一想，飞机沿跑道开始加速时，你会有什么样的感觉？你会感到一种与引力

非常相似的压力，将你推向椅背。如果加速度足够大，你就可以体验到宇航员或者战斗机飞行员经常感受到的“几倍重力”的感觉，你会觉得体重变得非常沉。

接下来，我们看看爱因斯坦当时想到了什么。宇航员和从高楼上掉下来的人在自由落体的过程中都感觉不到自身的重量，这个事实说明加速度与万有引力的体验完全相同。他在脑海里想象出一个与伽利略的设想非常相似的情境。当初，伽利略在考虑相对性时，想象自己身处一个平稳前进、没有窗户的船舱中。他发现，无论在这个船舱中做什么实验，都无法确定自己是在运动还是静止状态。爱因斯坦在研究狭义相对论时，还考虑了光速不变这个条件。现在，他又把加速度加了进来，要考虑的东西更多了。

假设你在一艘没有窗户的宇宙飞船中，有一个大小不变的力正在将你推向飞船的后部，使你感受到自己的体重。爱因斯坦称，飞船有可能正降落在某个星球上，且该星球的引力正好与将你向后推的力大小相等；飞船也有可能正在太空中，在引擎的推动下以恒定的加速度提高飞行速度。但是，你无法在飞船中通过做实验来判断它到底处于哪种状态。引力的拉动与加速度是等效的，二者无法区分。

爱卖弄的学究（喜欢卖弄的人往往与科学研究有密切关系）可能会辩解，严格地说，这两种情况是可以区分的，因为我们可以在飞船内部四处走动。既然如此，我们可以在飞船后部和前部各做一次实验。如果飞船在加速，两次实验结果应该没有什么不同，因为飞船各个位置的加速度是一样的。但是，如果是受引力作用，我们就可以从实验结果中看到细微的区别，因为飞船前部离星球更远，我们感受到的引力作用会弱一些。这个说法完全正确，但是它不符合等效原理，因为等效原理要求在太空中选择一个点，然后讨论该点的情况，而不允许我们在飞船中四处走动。

在这艘想象的飞船上，我们还可以做另外一个极其重要的实验。假设我们把狭义相对论实验中使用的光钟带上了飞船。我们知道，当飞船以稳定的速度飞行时，在飞船上的人看来，光钟没有移动，光束也不会受到任何影响，而是继续在两个镜面之间上下传播。现在，假设我们打开飞船的引擎，让飞船不再以恒定速度相对地球运动，而是开始加速。即使身处飞船内部，我们也能感受到飞船在加速。当光从顶部镜面反射到底部镜面时，飞船上的乘客可以看出光的传播路径发生了变化。

这个现象非常有趣，表明伽利略的“无法通过实验区分”相对性的断言在加速度条件下是不成立的。但是，我们无须光学实验就可以知道他的说法不对。我们的身体能感受到加速度的效应，我们还可以在手机上安装一个加速度计。但是，等效原理可以告诉我们一些更有趣的事情。我们已经知道，如果飞船加速运动，光的传播路径就会改变。既然如此，如果飞船在引力场中，肯定也会出现同样的现象。爱因斯坦由此意识到万有引力的一个基本特性：物质具有弯曲时空的能力。具有质量的物质可以使时空弯曲，因此，我们看到本来应该直线传播的光束的路径发生了变化。

这些话听上去似乎有点儿耳熟。还记得绕地球运转的空间站吗？空间站沿着地球切线的方向飞行，飞行路线也是直线。本来，空间站应该会飞离地球，进入太空，但是地球质量使时空发生了弯曲，于是空间站的飞行路线也随之弯曲，并形成了轨道。因此，当飞行物遇到巨大物体时，飞行路线会发生弯曲。但是，静止的物体（例如苹果）为什么会突然加速落到地面上呢？这里，我们需要了解一个重要的事实：发生弯曲的不是空间，而是时间。尽管苹果在空间中保持静止，但在时空中却不是静止的，因为时间一直在滴滴答答地流逝。也就是说，苹果加速掉落是时间弯曲的结果。

这些解释可以帮助我们了解广义相对论，但是目前还没有涉及相关数学知识。知道这些内容可以让我们对广义相对论有一个粗浅的了解，但还不足以让我们将它应用到科学研究中。爱因斯坦擅长运用有限的数学工具，把思想实验的情况直观地表现出来，然后添加大量的数学内容，明确这个思想实验可能产生的结果。在研究广义相对论时，添加数学内容的工作持续了几年时间，其中大部分工作是在1911年（因为对量子理论的研究让爱因斯坦痛苦不堪，所以他决定暂时停止这项研究）至1915年完成的。

爱因斯坦面临的第一个问题是，他需要摆脱欧几里得2 000年前深入讨论的几何学（参见第4章）。我们知道，古希腊人是在一个平整的表面上研究毕达哥拉斯定理与欧几里得几何定理的。他们认为这些都是理所当然的，以至于不愿意花费力气把它们表述成公理。但是，稍加考虑，我们就会发现这个假设条件有点儿奇怪。别的不说，单是真正平整的表面，古代就比现代少得多。柏拉图的完美原型自然可以做到真正平整，但是它们投射到现实世界的山洞中的阴影则不可避免有瑕疵。古希腊人规定了几何学研究的直线与现实世界不同，且没有粗细之分，但是没有任何人明确地表述过平整表面这个假设。

让我们感到奇怪的另外一个原因是，古希腊人知道地球不是平的，但他们还是一如既往地做出了相反的假设，也就是说，他们假设地球是平的。在中世纪之前，人们普遍认为平坦的地球是标准模型，但与此同时，所有受过教育的人都认为地球只能是一个球体，而且早在古希腊时代，人们就已经证明了这一点。因此，尽管几何学令古希腊人深感自豪，尽管它的名字包含“测量地球”之意，但是在研究地球的曲面时，这门科学其实并不适用。

前文说过，垂直于赤道的平行线相交于地球两极。这已经反映出几何学的问题了，如果我们再考虑三角形这个最常见的几何图形，就可以更加清楚地看出这个问题。欧几里得的《几何原本》（卷一）的

第32个命题告诉我们：“在任意三角形中，如果延长一边，则外角等于另两个内角之和，而且三角形的三个内角和等于两个直角的和。”换成我们熟悉的语言就是，欧几里得认为，三角形的三个内角和等于180度。

在欧几里得精心搭建的几何系统中（别忘了，在这个命题之前，还有31个其他命题），这个命题肯定是对的，不容置疑。可是，一旦离开了平整的表面，它就不成立了。在地球这个等球体上（地球至少是一个近似球体），与直线相对应的是大圆。显然，球面上的所有线不可能在所有三个维度上都是直线。事实上，所有线都是弯曲的，曲率与地球表面相同。大圆是指围绕地球表面且圆心与地球球心重合的所有曲线。

我们可以做一个简单的实验，利用大圆构建一个三角形，就可以彻底推翻欧几里得几何学。我们从赤道这个大圆上取两个不重合的点，然后从这两个点开始，向北极方向各作一条与赤道成90度角的直线（分别与另一个大圆重合）。这两条直线将在北极相交，并构成一个夹角。赤道上两点之间的距离越大，两条直线在北极形成的夹角就越大。接下来，我们计算这个三角形的内角和。赤道上的两个角各为90度，第三个角的度数还不确定，但它们的和肯定超过180度。事实上，如果赤道上两点之间的距离为10 000千米（约等于赤道至北极的距离。千米这个单位当时就是通过这个方法定义的），这个三角形就是一个等边三角形，三条边的长度都相等。此时，三个内角都是90度，因此这个三角形的内角和是270度。

由此我们知道，爱因斯坦在计算广义相对论中的时空曲率时，是无法使用欧几里得几何定理的，因为他使用的数学工具必须可以处理曲面。更令人吃惊的是，他甚至还要处理四个维度（包括三个空间维和一个时间维）全部弯曲的情况。爱因斯坦的朋友马塞尔·格罗斯曼为

他提供了一个解决方案：德国数学家伯恩哈德·黎曼创立的当时最先进的弯曲空间几何学。

此时，爱因斯坦的研究成果已经引起所有人，尤其是德国杰出的数学家戴维·希尔伯特的注意。希尔伯特曾邀请爱因斯坦到他任教的哥廷根大学做报告。之后，他对爱因斯坦的关注度进一步提高。当时，爱因斯坦的研究仍然有几个不完善的地方，希尔伯特似乎认为，他应该赶在爱因斯坦之前完成改进工作。事实上，尽管爱因斯坦先于希尔伯特发表了第一篇完整的广义相对论论文，但人们一度认为希尔伯特率先完成了那些方程，只不过他发表论文的时间晚于爱因斯坦。后来，人们发现希尔伯特的原始论文本身存在若干缺陷，他在看到爱因斯坦的论文之后才进行了修改，因此首发权仍然应该归属于爱因斯坦。

无论第一个完成这项工作的人是谁，第一个着手研究广义相对论的人毫无疑问是爱因斯坦。与牛顿和莱布尼茨在微积分开创者问题上的纷争不同，爱因斯坦与希尔伯特并没有相互指责，希尔伯特还大度地承认这是爱因斯坦的杰作。整个研究成果可以用一个看起来十分简单的方程表示：

$$G\mu\nu + \Lambda g\mu\nu = (8\pi G/c^4)T\mu\nu$$

除了光速的四次幂以外，整个方程看起来似乎非常简单，但是那些带有$\mu\nu$下标的参数都不是简单的变量，而是爱因斯坦从黎曼研究成果中借鉴而来的张量——一种功能强大但是难以操作的数学工具。张量可以把不同数值、矢量以及其他张量之间的多维关系压缩成一个单项。这个单项看似非常简单，但是其背后隐藏着一个变量矩阵。把引力场方程中的张量展开，将会变成10个基本微分方程，而且其中涉及的值会根据在空间和时间中的位置不同而发生变化。

牛顿的万有引力理论包含一个简单的基本方程（牛顿本人并没有使用这个方程）：

$$F = Gm_1m_2/r^2$$

整个方程只涉及常数G、两个天体的质量与它们之间的距离（r）。爱因斯坦的方程也有G，但是其他需要考虑的因素远不只是质量的影响作用，尽管质量非常重要。其中一个关键因素是，物体质量不固定且随着运动发生变化（根据狭义相对论）。能量可以增加质量。后来，爱因斯坦还证明了能量也可以增加引力。质量的效应则主要是使时间发生弯曲。

与此同时，空间也会发生弯曲。宇宙飞船加速时导致光束弯曲的根本原因就是空间弯曲。与时间不同，空间弯曲需要考虑所有三个维度，每个维度上有两个方向，因此爱因斯坦需要考虑的因素又多了六个。此外，他还要综合考虑一些稀奇古怪的东西，例如惯性系拖曳效应（由于相对性效应，运动物体会产生一个垂直于运动方向、强度不大的引力）。爱因斯坦需要考虑的最后一个问题是，引力会产生某种形式的能量（某些物体，如高山上的石头，因为位于引力场的高处而具有势能）。我们已经知道，能量可以产生引力。于是，它们就会形成一个小的反馈回路，引力又摇身一变成了一个引力源。

与狭义相对论不同，广义相对论使用的是寻常人并不熟悉，甚至永远也不会熟悉的数学知识。事实证明，这些复杂方程的求解极富挑战性。尽管在特例中解这些方程并非难事，但是它们没有一般性解法。麦克斯韦通过求解复杂方程，预测到电磁波这个意想不到的物理实体的存在，这是人们利用数学工具进行类似预测的早期实例。现在，根据广义相对论方程，人们又预测到黑洞的存在。

黑洞这个概念要追溯至18世纪。当时，英国天文学家约翰·米歇尔发现，如果天体的质量足够大，它的逃逸速度（摆脱该天体引力必须

具备的速度）就会非常快，甚至能超过光速，因此光也有可能无法从这种暗星上逃逸。后来，德国物理学家卡尔·史瓦西受到爱因斯坦相对论的影响，重新提出这个概念，那时米歇尔的研究早已被人们遗忘了。

史瓦西提出这个概念的时间是1916年，当时他正在参加第一次世界大战，而相对论也只发表了几个月。此时，爱因斯坦本人仅仅求得了部分近似解，做出水星轨道是变化的这个可以检验的预测。但是，躲在战壕里的史瓦西却得到了黑洞这个特例的精确解，指出星体耗尽支撑其重量的燃料之后，就会因为无法抵抗自身引力而坍缩，最终变成黑洞。

但是，史瓦西并没有把他的预测结果称作“黑洞”，这个名词直到20世纪60年代才出现。人们通常认为黑洞这个名称是约翰·惠勒发明的。虽然惠勒肯定是最早使用这个名称的科学家之一，但是第一个想出这个名词的人似乎不是他。1964年1月，在美国科学促进会的一次会议上，有人开始传播这个术语，但这个人不是惠勒。随后，安·尤因在《科学简讯》上发表了一篇介绍这次会议的文章，把这个术语变成了铅字。也就是说，这个名称源自那次会议，但是无法确定发明者。

无论黑洞这个名称源自何处，我们都知道黑洞从何而来，答案就是数学。在检验科学模型的有效性时，我们往往会考察它能否预测出当初没有被纳入模型的现实世界的情况。因此，爱因斯坦急匆匆地求得方程的部分近似解，并开始预测水星轨道的特点。不久之后，人们利用日食发生的时机，通过观察从太阳旁边经过的星光，证实了相对论关于引力可导致光束弯曲的预测是正确的。但是，在预测黑洞的存在之前，几乎没有人仅利用模型就预测出一个所有人见所未见，甚至想都没想过的物理实体。

即使是现在，与其说黑洞是科学研究的对象，还不如说它是数学的产物。天文学家在太空深处观察到很多天体，有间接证据表明，从

它们表现出来的特点看，这些天体似乎就是黑洞。尽管这些证据有很强的说服力，但所有证据都只是间接证据。我们从来没有直接观察到黑洞，只是借助数学工具推演出我们在银河系中心可能观察到的那些景象（据猜测，银河系中心有一个无比庞大的黑洞）。这里应用的数学绝对位于现实世界的边界线上，似乎与真实世界存在某种关系，但尚未得到证实。有人（例如马克斯·泰格马克）认为，数学与现实之间的联系永远达不到严丝合缝，如果我们可以近距离分析黑洞，就可以证明这些预测都是错误的。

尽管相对论的证明工作非常重要，但是在此期间，爱因斯坦还投身于物理学的一个新分支——量子物理领域，并在其中发挥了重要作用。量子物理是一个研究微观世界的物理分支，同样具有革命性意义。20世纪初，原子尚未被视为一种确定存在的实体，而只是一个有用的概念性工具，可用于预测物质有哪些特征。但随后，人们不仅证明了原子的确存在（主要是爱因斯坦的功劳），而且发现原子具有一些看似完全违背自然规律的特征。这实在是一个自相矛盾的悖论，因为自然界主要是由这种量子构成的。

量子物理非常复杂（它研究的是由原子、电子、光子等粒子构成的微观世界），本书不准备对其进行深入讨论，但是量子物理在其发展过程中得出了两个非常重要的观察结果，可以帮助我们了解数学对于宇宙探索的重要意义。爱因斯坦和丹麦物理学家尼尔斯·玻尔等人是最早进入量子物理领域的科学家。他们很快发现，如果从量子的层级来研究宇宙，就必须抛弃很多关于物质和光的特征的假设。例如，人们早就认为光是一种波，这个假设也已经得到了麦克斯韦的证实。但是，在新兴的量子物理研究领域脱颖而出的大咖们告诉人们，所谓的光波特性，不过就是一个模型。光的确具有类似于波的特性，但它也可以被描述成一堆粒子或者场扰动。

在解释原子中的电子与光的相互作用时，玻尔受行星绕太阳运行的启发，试图把这些电子放入类似的轨道。尽管这个模型早在20世纪20年代就已经过时了，但是我们仍然可以在几乎所有的原子平面结构图中看到它的身影。玻尔很快就发现自己的这个想法行不通，于是他为这些电子设计了多个轨道，让电子可以在不同的轨道之间跳跃，但是不允许它们停留在轨道中间。我们所熟悉的那些看得见、摸得着的“真实”事物不具有这种特征，但是这些量子却表现出这个特点。

玻尔的原子研究没有多少实用价值，但激起了一些年轻科学家的兴趣，其中最有名的当属沃纳·海森堡和埃尔温·薛定谔。他们决定在简单的玻尔原子模型的基础上，用数学描绘出量子的相互作用。海森堡更加彻底，他提出的矩阵力学利用纯粹的数学方法，描述量子的特点。他根本不考虑借助模型这种直观表示，而是通过操作矩阵（即数组），摇动手柄，使黑箱模型像麦克斯韦电磁波方程组一样，做出一些重要的预测。

薛定谔不喜欢这种抽象的方法，因此他利用波这种熟悉的形式，考虑如何建立量子行为的模型。时至今日，薛定谔方程对于我们理解量子物理仍然具有非常重要的意义。它同样属于形式十分简单、实质却十分复杂的方程：

$$i\hbar\frac{\partial}{\partial t}\Psi=\hat{H}\Psi$$

从数学的角度看，这个方程的某些特点足以让你胆战心惊。首先，该方程使用了–1的平方根i。其次，就像爱因斯坦方程中的张量一样，薛定谔方程中的希腊字母Ψ和戴帽子的H这两个符号都暗藏玄机。Ψ是描述某个系统本质的波函数，表达式非常复杂，而戴帽子的H指“哈密顿算符”，表示系统中的能量并将其应用到波函数之中。

最初，人们以为薛定谔方程中的波函数，或者说该函数值的平方（谢天谢地，终于摆脱了那个让人讨厌的i），表示的是量子在系统中

的位置。但是，真如此的话，似乎就可以预言所有量子无时无刻不在扩散，并且越变越大，但是没有实验可以证明这个预测。（谢天谢地，否则我们将身处一个无比奇怪的世界。）再次，人们发现该方程预测的是量子处于某个位置或者系统处于某种状态的概率。

除了这个方程和令人晕头转向的数学应用，量子理论研究还产出了大量其他成果。例如，将相对论引入量子行为研究的狄拉克方程（它有一个副产品，即预测了反物质的存在），以及在量子理论的基础上发展而成的量子电动力学等。但是，只需要看看海森堡和薛定谔，我们就可以看出科学家在“采撷”数学运算结果而不单纯依靠直接观察的道路上取得了两次重要的突破。

海森堡在他的矩阵力学中使用了完全数学化的描述方法，薛定谔紧随其后，他的方程把概率纳入其中。爱因斯坦辛辛苦苦地把量子理论带到我们这个世界上，但是在涉及概率之后，他发现在量子这个层级，如果不经过测量，所有的存在都只是一种可能性。例如，如果刚刚没有测量过，就无法确定量子的位置。人们常说，量子在同一时间里的可能位置有两个，但是真实情况似乎更加复杂，展现在人们眼前的是概率织成的一张网。这让爱因斯坦感觉很不舒服，因此他立刻停止了这项研究。

从本质上讲，量子物理似乎从根本上把现实变成了数学。根据量子理论，如果不是正在测量，关于现实世界的所有观察结果对概率的依赖程度都会很大。乍一看，这与抛硬币似乎没有本质上的不同。在公平的抛硬币游戏中，得到正面与反面的概率通常各占一半。但是，一旦硬币被抛出去，出现某种结果的可能性就是100%，而出现另一种结果的可能性则是0。只不过在硬币落地之前，我们不知道这个结果到底是什么。但是，如果抛出去的是一枚量子硬币，我们得到的就真的只有各占一半的概率，在进行测量并得出结果之前，前面说的那种潜在确定性是不存在的。

尽管数学概念与真实的量子之间的这种关系似乎不可信，但是多年来已经有大量实验证明了它的真实性。事实证明，“现实世界建立在数学基础之上”的说法属于夸大其词。从某种意义上讲，这个说法已经不像以前那样令人吃惊了。数学仍然是现实的模型，而不是一种绝对的描述。概率与推动现代物理发展的抽象数学不同，它是作为应用数学工具被人们发明出来的。概率源于抛硬币等真实操作的观察结果，只不过在量子物理中，它表现出之前没有的独立性。这并不是说量子就是概率，而是说我们通常只能以概率的形式来描述它们。

即使在伽利略和牛顿的时代，物理学领域也不时可以看到对称的身影，例如牛顿第三运动定律就描述了一种令人赏心悦目的对称，人们对相对论莫衷一是的态度也形成了一种对称。但是，随着历史的车轮进入20世纪并继续前行，对称展示出新的活力，对科学理论的发展起到显著的推动作用。从此，数学真正地占据了主导地位。

第14章
诺特：对称之美与隐形恶龙

Are Numbers Real?

科学界和数学界涌现出的伟大女性屈指可数，这不是因为她们在这两个领域的成就比不上男性，而是因为长期以来人们一直认为这两个领域极具挑战性，不适合女性。在这种奇怪观念的影响下，女性被科学和数学拒之门外。如果问20世纪上半叶及之前有哪些女性在数学或科学领域取得过伟大的成就，任何人都有可能提到玛丽·居里、埃达·洛夫莱斯。如果继续追问，他们可能还会提到卡罗琳·赫舍尔，但是他们不大可能说出艾米·诺特。然而，这位德国女数学家仅凭一己之力，就创立了抽象代数理论。具体来讲，诺特明确了对称的重要性。她发现，不仅自然界有大量的对称结构，在很多物理法则的背后也有数学对称的身影。

如今，在数学的指引下，寻找现实中的匹配对象的研究并不少见。诺特发现对称可以推导出能量守恒定律，她的发现使这个由麦克斯韦开创的方法趋于完美。我们借助模型研究构成宇宙万物的基本粒子，但是这些标准模型并非源于那些别出心裁的实验，例如大型强子对撞机，而是得益于诺特及其追随者的数学推演。事实证明，这种方法在很多方面都取得了显著的成绩，但是有时候你又会觉得整个推演过程似乎走得太远了。

如果说导致艾米·诺特名声不那么显赫的唯一原因在于她是一位女性，也许会显得有些偏激。但是，如果你请公众随便说出几个20世纪的数学家，他们说出的人可能都是科学家，而不是数学家。实际上，从诺特近期的受关注程度来看，人们很有可能会提到她的名字。但是，鉴于她的研究对物理学的发展具有非常重要的意义，我们有理由相信她的名气应该不亚于薛定谔与海森堡。

1882年，诺特出生于巴伐利亚，她的父亲是受人尊敬的数学家马克斯·诺特。父母给她起的名字是艾米莉，但是年幼的诺特喜欢称呼自己艾米，并很快让大家接受了她的这个新名字。与很多大数学家、大科学家不同的是，诺特上学时对数学并不感兴趣。她先是取得了语言教师资格，1903年她考入埃尔朗根大学，并于1907年成为德国有史以来第二位女数学博士。

1909年，诺特在爱因斯坦的挑战者希尔伯特的建议下，来到哥廷根大学。1915年，希尔伯特提议学校授予诺特特许任教资格（在德国申请任教资格需要满足一个独特的条件，即写一篇类似于博士后论文的文章）。但是这个资格只对男性开放，因此希尔伯特及其支持者向政府请愿，希望他们特事特办。最初的请愿遭到了拒绝，直到1919年诺特才获得了特许任教资格，但她始终没有成为一名教授。同众多数学家一样，她最重要的研究成果都是在年轻时完成的。诺特定理（我们马上就会讨论）是她刚到哥廷根大学的时候提出来的。诺特还有一些非常重要的数学发现，但是一旦超出纯粹的数学圈子，这些发现的影响力就比不上诺特定理了。

1933年，刚刚掌权的纳粹剥夺了诺特的任教资格，终结了她相对平静的教学生活。人们通常认为这件事主要归因于她的犹太血统，但这可能只是其中一个原因。诺特与几名不是犹太人却同样遭到驱逐的同事有一个共同点，那就是她也同情共产党，并且于1928年接受了莫斯科一所大学的客座教授称号。同莫斯科的这种联系更有可能起到导

火索的作用，因为她的命运很快就变得坎坷起来。诺特试图回到莫斯科，但是事实证明，在那个动荡不安的年代，她的这个想法很难实现。于是，诺特搬到了美国。1935年，她在美国去世，年仅53岁。

要理解诺特定理及其对理论物理学家的重要性，我们就必须知道什么是对称，对称为什么会产生这些重要的结果，这种方法对数学和现实之间的关系有哪些拓展作用。在日常英语中，我们所说的对称是指反射对称，也就是镜像对称。很多生物身上都存在这种对称。如果有机体有“左”、“右”两侧，我们就知道它们是镜像对称。实际上，这种对称性在人类对美的理解中占十分重要的地位。

对美的判断标准可能因人而异，但是我们通常都会根据某些特点来判断一个人的外表是否美丽，对称是其中最重要的标准之一。多次测试的结果表明，人类通常觉得对称的脸比不对称的脸更吸引人（奇怪的是，鸡也有同样的特点）。人们认为，这可能是因为显著的不对称往往是疾病导致的。如果美的主要作用是帮助吸引潜在配偶，缺失对称性就可能意味着生殖能力的降低。对对称美的追求因此成为一种进化特征。

除了简单的镜像对称，数学家还发现了各种各样的对称。一般而言，如果某个物体发生某种变化之后没有出现可以辨别的不同，我们就说这个物体是对称的。比如，对于左右镜像对称，我们可以将镜像的左右两边对调，得到的镜像没有任何变化。借助简单图形，我们可以更方便地理解对称。将矩形或正方形的左半边与右半边对调，将会得到无法区分的镜像结果。

旋转对称是另外一种简单的对称。将正方形旋转90度后得到的图形与原来的正方形一模一样，没有任何不同（矩形则不适合这个操作）。但是，如果将正方形旋转45度，情况就不一样了。我们可以看出发生了某种变化，因为旋转后的图形与之前不同了。圆与正方形不

同。圆的旋转对称性是最强的。将圆旋转任意角度，哪怕是不到一度，得到的圆都与原来一模一样。

除了这些简单的对称，其他对称需要稍微拓宽对称性的应用范围。例如，我们可以考虑时间上的对称。如果经过一段时间之后，我们看到的情景没有发生变化，我们就可以认为在那段时间里该情景关于时间对称。静止物体看起来总是不变，因此它们在时间上的对称性最强。发生周期性变化的物体（例如钟表的分针）在时间上具有有限的对称性，这一点与正方形的旋转对称性比较相似。

在研究时间上的对称性时，我们从某些非常明显的例子中看出相对性会影响对称。我们可以说完全静止的物体对称性最强，但是“静止”是一个相对概念，而不是绝对概念。例如，我们考虑一下宇宙飞船在时间上的对称性。从飞船乘客的视角来看，飞船是静止的，因此它的对称性最强。在他们看来，飞船的外表（仅考虑飞船的外壳，而不考虑任何磨损情况）没有发生任何变化。但是，飞船外面的人看到飞船一掠而过，时间上的对称性就遭到破坏。

另外一种对称——平移对称，只有数学家才会觉得它理所当然。在研究平移对称时，我们不会考虑时间，而是比较两张“照片”，看看它们有没有不同之处。例如，如果将正方形向右方移动一个边长的距离，那么在现实世界中，我能看出这个正方形的位置发生了变化，因此这个操作显然不具有对称性。但数学家却在脑海中想象这个正方形通过不断重复的方式形成一个无限延伸的表面，正方形本身就是这个表面的一部分。然后，他们问道：在这种情况下，发生这种移动后你能看出它有任何不同吗？答案是“不能”。也就是说，纯粹的数学世界中存在平移对称。然而，如果正方形移动的距离不等于边长，而是小于边长，最后得到的图形就有了变化，对称将不复存在。

毫无疑问，自然界中近似对称的现象比比皆是。很多物理结构都近似对称，这并不是因为对称有某种魔力，而是因为它是一种效率很

高的发展方式，或者因为形成这种结构的作用力在各个方向上都相同。大多数动物都具有某种对称性，例如，我们已经讨论过的左右对称，或者像海星那样更加吸引人的旋转对称。雨滴、鸡蛋、草的叶子、星体都表现出对称性。这些事物在现实世界中的对称性可能并不完美，但至少它们投射在柏拉图的山洞中的影子是对称的。在对称性上能达到这种程度的唯一“实物”可能就是黑洞的事件视界吧（假设黑洞真实存在）？

此外，物理学中也有很多对称，除了牛顿第三定律的反射对称——“每个作用力都有一个大小相等、方向相反的反作用力”，我们还发现某些法则在时间、空间上发生位移或者旋转后仍然适用。如果我们转过身，面朝另一个方向时，力与加速度之间的关系就会发生变化，我们这个世界将十分精彩，但是这种不一致性会导致我们很难有效地开展科学研究。

同理，我们必须假设物理定律不会因为时间或位置的变化而改变，否则物理研究就无法进行下去。值得庆幸的是，迄今为止的经验告诉我们这个假设可能是正确的，尽管我们可能永远无法给出证明。如果重力不断变化，那么所有涉及重力的物理学都将失去意义。几乎所有科学模型的出发点都是一条公理：物理定律和常数都具有对称性。但是，这条公理纯粹是一种基于实用主义的判断，也几乎是所有科学模型的出发点。比如，如果光速不断变化，宇宙学和天文学就会手忙脚乱，因为在观察远处的天体时，我们需要在光速恒定的基础上推算我们看到的是多久之前的情景。

偶尔也有一些习惯于反传统的人会对这条公理指手画脚。比如，物理学家安德里斯·阿尔布雷切特与若奥·马奎荷就曾指出，自宇宙诞生以来，光速已经发生了显著变化，因此不需要引入宇宙膨胀这个概念。（宇宙膨胀说认为，宇宙肯定在一段时间里发生过突如其来且原因不明的大幅度膨胀，从人们推断的诞生之日起就因为距离过远而彼

此从无接触的宇宙的各个部分，通过这个膨胀过程达到了均衡状态。）但是，在大多数时间里，没有人会质疑这条公理，因为如果没有这个假设，我们几乎不可能进行物理研究。

这个方法与科学家们口中流传的那个关于找钥匙的古老笑话有点儿相似。丢钥匙的那个人知道在他走到家门口的那条大街之前，钥匙就已经丢了。但是，他却一直在那条大街上寻觅。朋友问他："你为什么要这么做呢？你肯定找不到钥匙的。"答案很简单，因为在他回家的路上，只有这条大街上有路灯。其他地方都是黑漆漆的，即使去那些地方寻找也毫无意义。同样道理，尽管假设物理定律与常数不会随着时间、空间的变化而改变的做法可能大错特错，但是离开这条公理的"灯光"，试图从事科学研究的努力也将毫无意义。

虽然如此，我们还是找到了一些证据，至少可以证明某些物理定律与常数在时间上具有对称性，从而增加了这个假设的可信度。证据之一就是电荷的大小。借助1972年在西非加蓬奥克劳发现的一些引人注目的天然核反应堆，我们可以证明至少20亿年以来，电荷大小在时间上具有对称性。当这些天然反应堆形成的时候，铀238的含量远远多于裂变反应堆所需要的量。由于含量充足，所以当这些同位素在奥克劳聚集起来之后，核链式反应就开始了，并产生了大量的热和辐射。

实际上，如果钐的原子核质量稍有不同，这些核反应堆就不会形成钐这种产物。但是，我们知道，由于质能等效，所以原子核的质量取决于原子核中的质子携带的电荷。根据在这些年代遥远的反应堆中发现的钐，我们可以推断现在测量的质子电荷与当时的质子电荷相比，相差不到千万分之一，否则，这些钐根本就不会存在。诸如此类的测量结果至少可以确定某些常用的物理定律与常数在几十亿年时间里几乎不会发生变化。有必要指出，即使存在差异，也可能只发生在宇宙刚刚形成的时候，也就是说，在130多亿年（其实这个数字也是在相关假设的基础上推算出来的）前，而不是20亿年前。

诺特定理具有突破性意义，证明了在对称与能量守恒定律之间存在一种不可分割的联系。比如，如果实践证明某些物理定律不会随着时间的推移而发生改变，能量就必然在一个封闭系统中守恒。反之，如果我们发现能量是守恒的，这些物理定律就必然在时间上具有对称性。这个特点同样适用于其他对称性。如果旋转之后，系统的表现没有任何变化，那么角动量必然是守恒的。如果空间位移不会导致变化，就说明线动量守恒。

在此之前，能量守恒定律在某种程度上一直是以假设的形式存在的。对称不仅可以帮助我们理解其中的原因，在量子电动力学等量子场论显著改变了微观物理学之后，对称还为我们选择新的研究方向创造了条件。此外，在对称性的帮助下，构成宇宙万物的粒子的复杂本质进一步展现在我们眼前。

对称与量子层级出现的那些令人困惑、变化不定、随机的混乱局面到底有什么关系呢？物理学家需要化繁为简，才能理解现实，进而构建出现实世界的模型。对称为物理学家指了一条明路，尽管走上这条道路必然会遇到难以处理的数学问题。人们在原子核中发现中子的存在后，就踏上了这条道路。欧内斯特·卢瑟福曾预言，大多数的原子核中都存在一种不带电的粒子，即中子。之后不久，英国物理学家詹姆斯·查德威克证明这种粒子确实存在，并指出使质子结合在一起的那种未知作用力（现在叫作强核力）也会作用在中子身上。

质子带正电荷，当它们聚集在一起时，就会因为携带同种电荷而相互排斥。它们在原子核中结合得非常紧密，相互之间的排斥力肯定非常大，这说明将它们结合到一起的力更大。根据查德威克的发现，这种作用力似乎还会吸引不带电荷的中子。在中子被发现之前，人们并没有发现这个问题。如果原子核中只包含质子，就表明还存在电量相等的异性电荷，而且同性电荷被吸附在一起。但是，这种力还对中

子有效，情况就大不相同了。这时候，有一个人站了出来，将对称引入原子核行为研究，并赋予了它核心地位。这个人就是沃纳·海森堡。

我们在上一章已经讨论过，海森堡用来研究量子相互作用的数学模型在宏观世界找不到对应物，我们也无法借助现实世界的类比来理解这个模型的作用原理，但这个模型没有导致任何问题，而且矩阵中的数字与现实世界的观察结果很吻合。海森堡认为，这个模型已经非常完美，不需要添加任何内容了。接下来，海森堡做出了一个非常大胆的举动：通过一种其实并不存在的对称性来研究原子核。但是，他使用的是“类似对称”的说法，似乎在暗示有某种重要事物隐藏在背后，但是他没有说明。

海森堡发现质子和中子的质量非常接近，两者之间只存在大约0.14%的差别。此外，原子核中质子和中子的数量往往相仿。不可否认，质子和中子数量不等的原子核也非常多。例如，最简单的原子——氢原子只有一个质子、没有中子，而简单元素锂的主要同位素锂–7有3个质子和4个中子。但是，一般而言，在最稳定的原子核中，质子和中子的数量几乎相等。

在对称思想（而不是实际数据）的启发下，海森堡似乎很自然地认为，将原子核中的粒子结合到一起的力，对于质子和中子应该具有相同的效果。海森堡还认为，如果将原子（包含相同数量的质子和中子）中的中子换成质子，或者将质子换成中子，原子核的状况应该也不会发生任何变化。也就是说，原子核似乎表现出某种对称性。海森堡把这种对称性称作同位旋。除了容易导致混淆以外，我们看不出这个命名方法有什么明显的原因，因为这个概念与旋转没有任何关系。海森堡假设质子和中子的同位旋大小相等，性质相反（质子的同位旋为+1/2，中子的同位旋为–1/2），结果发现根据这个假设得出的预测与现实正好吻合。

随着各种各样的其他粒子被人们发现，美国人默里·盖尔曼提出了粒子在另一个方面表现出的对称性，并称之为奇异数，从而拓展了同位旋的概念。根据他的研究结果，这些粒子按照奇异数和同位旋的不同形成了整齐的排列。如果绘制成图，就会发现每8个粒子结合在一起，表现出一种引人注目的对称性。这种对称性表明，有某种根本原因促使它们形成这种结构。也就是说，它们之所以表现出这种对称，是因为它们的结合具有某些特点。

然而，尽管按照假设的同位旋与奇异数，可以看出这些粒子表现出明显的对称性，但仍然存在一个问题：所有这些都建立在越来越疯狂的推理的基础上。它们的对称不是很完美，比如，中子和质子的质量并不相等。等到其他粒子（例如"级联粒子"）被引入之后，情况就更糟糕了。级联粒子同中子和质子属于同一类别，质量比其他粒子大40%。即使真的有对称性，也会被这些粒子破坏殆尽。然而，对称这个数学概念似乎有极强的诱惑力，让人难以放手。

盖尔曼努力寻找可以让粒子友好共处并且形成这种对称性的基本结构。他发现，最明显的原因就是质子、中子等假定的基本粒子是由更小的成分构成的，而且同位旋与奇异性就是由这种结构造成的。质子、中子以及从宇宙射线和粒子加速器中发现的大量新粒子似乎与原子一样，也是由更小的粒子构成的。这些粒子齐心协力，创造出我们观察到的或者假设的大小不等的电荷、同位旋和奇异数。

现在，我们把那些更小的组成部分叫作夸克。这个名字是盖尔曼取的，人们通常认为他是从詹姆斯·乔伊斯的"小说"《芬尼根守灵夜》中获得的灵感，因为书中有"对麦克老大三呼夸克"这样的诗句。但是，盖尔曼却说他起名时首先想到的是"quork"（郭克）这个词，然后才联想到那句诗，并且接受了人们普遍把这个名称错拼写成"quark"（夸克）的事实。不管怎么说，这个名字也比"埃斯"（ace）

强。埃斯是欧洲核子研究中心的乔治·茨威格为他独立提出的一个类似概念取的名字。最终，“夸克”这个名字沿用下来。

光子最初是普朗克为满足数学工具的需要而提出的一个理论概念，与之相似，盖尔曼提出夸克这个概念时，也是为了简化数学结构，他并不认为真的存在这种粒子。但是现在，人们不仅认为这种粒子确实存在，而且认为它们（可能）是一种基本粒子。按照盖尔曼最初的设想，夸克有三种“味”，即上夸克、下夸克和奇夸克（由此可以看出它与乔伊斯的那句诗的确是有关系的）。将两个或三个夸克结合在一起，就能构成我们观察到的所有粒子。此外，如果构成质子和中子的上夸克和下夸克在质量上略有不同，而且奇夸克的质量略大一些，就可以解释为什么粒子间的对称性会遭到破坏了。可观察粒子的对称性确实存在，但这不是说它们的质量具有对称性，而是说构成这些粒子的基本粒子具有对称性。

事实上，仅有三种味是不够的。随着观察的深入，局面似乎越发混乱。大约20年之后，量子色动力学产生。盖尔曼受到描述光与物质相互作用的量子电动力学理论所取得的成绩的启发，选择了这个相似的名称。量子色动力学认为，夸克之间的力是由胶子这种粒子产生的，它不是一种简单的力，而是会呈现出各种各样的“颜色”（夸克的味不同于现实世界的味道，同样，夸克的色与现实世界的颜色也没有任何相同点），包括红、绿、蓝。事实上，夸克一共只有三种颜色，夸克的反粒子则呈现互补色，即反红色、反绿色和反蓝色。

这个系统的高明之处在于，这些夸克结合在一起之后一定会呈现白色。因此，在由三个夸克构成的粒子中，例如质子或中子，这三个夸克必须分别是红色、绿色和蓝色，而由两个夸克构成的粒子，如介子，则必须包含某种颜色和它的互补色。要让整个系统正常运转，还必须有与颜色有关的胶子，一共有8种。从数学的角度看，这个方法了不起的一点是：它围绕这种“色荷”重新形成对称。由于质量不同，夸

克的对称性不可避免地具有不完美的特点，而没有质量的胶子却表现出完美的对称性。

随着理论物理学家和应用数学家的研究不断深入，他们发现对称似乎可以揭示宇宙中更多的深层次奥秘。对称似乎是一种自然趋势，也就是说，人们越来越倾向于用对称这个数学工具去推导现实的本质。但是，这样做存在一个问题。我们知道粒子的质量并不真的具有对称性，我们还知道自然界的各种力也大相径庭。即使真的存在某种深层次的对称性，似乎也早已被破坏殆尽了。从人们新近用来研究宇宙起源的模型来看，宇宙在刚开始的时候是完全对称的，但是这种对称性早已不存在。在这个过程中，到底发生了什么？

这里存在一个与集合论选择公理有几分相似的问题。在人类干预的前提下，从集合中选择一个元素是一件易如反掌的事情，但是在没有人参与时，该如何选择呢？同样道理，如果对称一度存在而随后遭到破坏，那么导致对称破缺的原因到底是什么？人们需要找出"对称自发性破缺"的机理所在。人们经常举铅笔的例子来说明这个问题。将铅笔的笔尖朝下，直立在桌面上。铅笔必然会倒下，破坏直立的平衡状态，然后笔尖指向某个方向。但是，我们无法预测铅笔会在直立的状态下倒向何方。

不幸的是，这个例子是有缺陷的。如果铅笔真的处于完美的对称状态，就绝不会倒下，因为它必须受到某个力的作用，才会倾倒。就像电影《盗梦空间》中的那个陀螺一样，铅笔会一直保持明显违背自然常识的直立状态。只要对称遭到一点儿破坏，例如铅笔的平衡或笔尖的形状有瑕疵，或者受到气流等外力作用，铅笔就会倒下。我们生活在一个不对称的世界中，所以我们知道铅笔肯定会倒。

对称被广泛地应用于探索宇宙奥秘的活动中，已经成为人们提出物理理论的主要手段。但是，我们必须十分小心。诺贝尔物理学奖得

主利昂·莱德曼认为，如果我们假定某些对称性真实存在，我们构建的科学模型就可能具有误导性。他说：

> 对称，即使是现实世界中的近似对称，也可能是一个功能强大的工具。但是，我们人类经常犯错，认为某些事物表现出完美的对称性。实际上，这些对称只是人们的错觉，或者是其他事物不经意间造成的偶然结果。

尽管这个例子并不完美，但是它确实可以说明近似对称的自发性破缺。人们认为，我们现在已经发现的各种力就是源于这种自发性破缺。当系统由高能状态进入低能状态时，就会发生这个过程。高能状态更有可能具有随机性，因此表现出更明显的对称性。与倒在桌子上的铅笔相比，直立的铅笔具有更多的势能。与之类似，加热传统的磁体并使它超过某个温度水平，它就会失去磁性。这是因为在热动能的作用下，磁体内部的磁畴由整齐排列变成了杂乱无序的随机排列。借助数学模型，我们可以将弱核力和电磁力统一起来。在宇宙诞生之初，弱核力和电磁力似乎是统一的，但是，随着宇宙的温度不断降低，对称自发性破缺最终导致这两种力的分离。

尽管数学模型非常简洁，但是它与我们观察到的现实并不完全吻合。对称理论要求负责传递作用力的玻色子（例如，传递电磁力的光子和传递强核力的胶子）必须没有质量。光子和胶子的确没有质量，但是负责传递第三种力（核裂变时出现的、可以使一种粒子变成另一种粒子的弱核力）的三种粒子都有质量。于是，对称理论似乎被彻底颠覆了。

对于某些人而言，数学是推动物理发展的全部动力，而对称理论的这些瑕疵令人无法接受，因此他们决定想办法做出补救，以便对称可以继续发挥推动作用。于是，他们提出了一个非常大胆的想法：有没有可能这些携带弱核力的粒子真的没有质量，但是宇宙中却存在一

个力场，就像电磁场（以及其他场）那样充斥在宇宙中，为自然界提供另外一种作用力呢？这个力场非常特别，它的唯一作用就是拖曳粒子，使携带弱核力的玻色子产生具有质量的假象。人们以该理论的一个创立者的姓名将其命名为“希格斯场”。

在用希格斯场打好“补丁”之后，人们发现整个系统中可能还隐藏着一种对称性。唯一的问题是，没有证据可以证明希格斯场的确存在。希格斯场是人们弥补理论瑕疵的应急之举，具有主观随意性，没有得到实验数据的支持。因此，寻找希格斯力的携带粒子——希格斯玻色子的工作具有非常重要的意义。2013年，欧洲核子研究中心的大型强子对撞机实验得出的一些结果在令新闻媒体摸不着头脑的同时，又让这些媒体欣喜若狂，原因是这些实验结果与希格斯玻色子存在的假设并不冲突。但有必要强调一点，所有这些实验结果都是间接证据，而且这套理论无法预言希格斯玻色子的质量到底是多少。

目前，人们利用数学推导结果搭建而成的模型，在诸多领域取得了实实在在的成绩。例如，在上述对称理论基础上建立的粒子物理标准模型就是一个成功的案例。尽管还有若干问题有待解决，但是它的很多预测与现实高度吻合。然而，这个模型中有很多要素都来自于直接的观察结果，而不是根据主体结构的预测。到目前为止，该模型还无法解释宇宙中暗物质（人们认为，暗物质的数量多于普通物质）的本质，也无法解释对称性和质量的存在原因。我们只知道某些事物表现出某种对称性，或者具有质量，但是无法解释具体的原因。此外，在物质粒子与携力粒子之间也没有发现明显的相关性。

为了回避其中的某些问题，人们再一次求助于数学工具。在成功应用对称理论的经验的引领下，有人提出了一种叫作超对称的全新对称概念，将这两大类粒子联系到一起。这个理论唯一需要解决的就是“简单化”问题。由于每种粒子都需要找到一种与之相反的超对称粒子，以致标准模型过于复杂。光子、胶子等携力玻色子需要与光微

子、胶微子等物质粒子对称，与此同时，电子、夸克等构成物质的费米子也需要有与之对称的携力粒子，即所谓的超电子、超夸克。

在我创作本书的时候，人们还没有找到任何关于超对称粒子存在的证据。如果这套理论是正确的（目前还没有任何合理的理由可以证明这套理论是正确的），在完美对称的宇宙之中，粒子和它们的超对称粒子就应该具有相同的质量，我们也就可以轻松地探测到超电子的存在。由于至今还没有探测到这些超对称粒子，所以我们必须彻底打破这种对称性，使这些超对称粒子的质量增加至等于或者大于希格斯玻色子的质量。也就是说，如果用大型强子对撞机完成更高能量水平的实验（这是未来的目标之一），就应该可以增加超对称理论发挥作用的可能性，但更有可能会起到反作用。

然而，对于利用纯粹的数学方法推导出物理理论的“魔术表演”而言，超对称只是一道开胃菜，而在“简单”的超对称理论中加入大量全新内容的弦论才是精心烹制的大菜。从弦论的总体描述来看，这套理论似乎兼具简单、美观这两个特点，但是一旦进行深入研究，就会发现它的复杂程度非常高，难点体现在具体内容上。总的来说，弦论就是用弦这个单一的一维实体来取代那些纷繁复杂的基本粒子。

就像电子不是小球一样，弦论中的弦显然也不是现实世界中的弦。但是，人们可以通过想象，让这条看似简单的基本实体以不同的方式（例如开弦或者闭弦）发生振动，从而发现所有可观测的粒子（包括物质粒子和携力粒子）。理论学家将抽象数学层层叠加到弦论之上，直到可以自圆其说，但也付出了高昂的代价。刚开始的时候，弦论存在5个主要版本，彼此之间水火不容，但是最后人们将它们统一起来，建立了M理论。

然而，弦论（或M理论）面临着严峻的考验。其中难度较小的一个考验是，这些理论不适用于包含三个空间维和一个时间维的传统概念，因为它们要求必须有9个（弦论）或10个（M理论）空间维。显

然，我们无法看见这些维度。因此，还需要有“补救措施”。于是，它们假设这些看不见的维度都蜷缩成非常小的一团，虽然我们看不见，但是它们依然可以发挥各自的作用。

弦论面临的一个更大的问题是，它会产生数不清的可能结果。弦论给出的可能解甚至比宇宙中的质子数量还多。结果的数量之多超乎人们的想象，数学却不会告诉我们如何取舍。物理学家马丁·波乔瓦尔德指出，弦论包罗万象，毫无疑问是一种万能理论。而且，弦论没有给出任何可以检测的预言，与现实没有任何联系。

英国物理学家保罗·戴维斯说：“（由于复杂程度非常高，又缺少预言，因此）研究弦论和M理论的学者们在检验真实性方面都有所欠缺。所有人最后得到的只是猜测。也许，他们在误打误撞之下，闯进了科学圣殿。果真如此的话，说不定哪一天他们就会告诉我们弦论的作用原理。又或者他们从此远离尘世，躲入桃源秘境。”有人认为，他们过于依赖抽象数学。如果利用数学推演理论，最后得到的都是虚无缥缈的空想，那又有什么意义呢?

数学不一定要与物质世界有相似之处，这不是它与生俱来的使命。原始社会的人费力地掰手指计算山羊或者玉米的数量时，可能会与现实世界形成直接的联系，但是人们很快就发现，负数及其平方根并不存在于我们周围的世界中。然而，即使引入这些量，也没有使数学失去意义。对于研究纯粹数学的人而言，与现实世界脱节不会导致任何问题。例如，纽结理论中的绳结与我们在现实世界中看到的任何绳结都不相同。在拓扑学中，甜甜圈与茶杯并无区别，重要的是如何迎接挑战，证明相互关系，推导出结论。只要不把茶水倒在甜甜圈上，就不会有任何问题。

在某种程度上，这种自由性可以发挥巨大的威力。在纯粹的数学研究中，现实世界的所有限制条件都无须考虑。不喜欢2 + 2 = 4，是吗？觉得有点儿厌烦？那么，我们可以让2 + 2 = 5，然后看看有什么

结果。这个等式在数橘子时可能不成立，但在数学世界里却是完全有可能成立的。与之类似，数学家早就认为三个空间维和一个时间维的限制具有主观随意性。他们发现，把研究工作搬到“相空间”之中，进展往往会非常顺利。就像物体有很多可能的状态一样，相空间的维数也非常多，甚至可以达到上万亿的数量级。这些维度并不存在于现实世界，但是在数学中却可以发挥重要作用。尽管弦论需要用到9或10个空间维度，尽管从技术角度看这些空间维度似乎根本不存在，但在数学家将物理学推向弦论这个研究方向时，这些问题似乎无关紧要。

说到这里，该哲学家卡尔·波普尔登场了。在当今科学界的权威人士眼中，波普尔似乎已经过时了。这是因为波普尔对科学本质的看法非常极端，公开宣称在科学研究中不应该使用归纳推理的方法。我们在根据已有的不完整观察结果做出各种预测时，使用的工具就是归纳推理。例如，我们推断光速是恒定的，因为根据我们的长期观察，光的传播速度都是恒定的。波普尔称，这个理由并不充分，因为说不定明天光速就会发生变化。没有归纳推理的话，科研几乎寸步难行。因此，波普尔的这个观点显然不切实际。但是，这并不意味着根据他的另一个观点总结出来的简化理论不可靠。

波普尔认为，任何科学理论都必须可以通过观察予以驳斥。反对波普尔的人通常认为，如果科学理论被任何观察证伪，我们通常就会拒绝接受这些理论，因此我们在应用波普尔的这个研究成果时必须非常谨慎，知道在什么时候、什么情况下才可以摒弃某个理论。波普尔的证伪机制在应用时显然要受到某些限制。在提出正式的质疑之前，我们必须认真检查，还需要重复证伪的过程。尽管如此，我们仍然有足够的理由断言，我们还是需要对付所谓的“隐形恶龙”，而且很多依据纯粹数学推导得出的现代物理学理论都难逃这个命运。

从隐形恶龙说可以看出科学必须接受波普尔原则的检验。设想某人说“我的车库里有隐形恶龙”，然后请科学家核实这句话。科学家肯

定看不见这些恶龙，触摸的方法也不能用，因为触摸恶龙会有危险。此外，恶龙可以飞翔，躲开科学家的触摸。也许科学家可以把面粉撒到地上，以便寻找恶龙的脚印。但是，这个人说，他车库里的恶龙非常特殊，都没有质量，也就是说，它们不会留下脚印。于是，科学家打算使用红外探测仪。但是，这个人又说，恶龙有非常好的热绝缘性。于是，科学家又打算探测恶龙移动时引起的空气振动，但是这个人说恶龙都是通过量子隧穿移动的，根本不会扰动空气。

由于没有办法观测到这些隐形恶龙，所以用科学的手段不可能发现这些恶龙。但是，这并不意味着这些恶龙不存在。恶龙有可能真的存在，而且它们有能力逃避所有的侦测手段。我们既不可能通过观察证明这些恶龙不存在，也没有办法对其进行科学研究。即使这个人可以创建出简明有效的数学工具，并证明他的车库里应该有隐形恶龙，情况也不会改观。

也许我们会感到奇怪，为什么波普尔要求理论可以被驳斥呢？在上面的例子中，那个人希望证明自己家的车库里有恶龙。但是，只有反面证据才具有确定性。即使在车库里找到爪痕，也不能确定它们就是恶龙留下的。那个人可以趁科学家不在场的时候，伪造出这些爪痕。但是，如果恶龙存在说提出了一个可以验证的预言，验证结果表明这个预言绝对是错误的，我们就可以判断恶龙存在说是不正确的。也就是说，除非结合失败的预言对理论进行修改，否则这个理论就是不正确的。真正的科学经常会遭遇预言失败的情况，因此，波普尔的这个方法在实际应用中有时会遭遇困难。然而，这些困难无法抹杀它的价值。

科学常常无法提供证据。“事实”不足为凭，我们需要的是“证据确凿的理论”。黑天鹅就是人们经常援引的一个例子。在几百年时间里，欧洲人一直认为世界上只有白天鹅，因为人们观察到的所有天鹅都是白色的，也就是说，所有证据都支持“所有天鹅都是白色的”这个说

法。但是，这个证明并不科学。的确，我们看到的所有天鹅都是白色的，但是这个事实无法证明所有的天鹅都是白色的。如果有人从澳大利亚带回一只黑天鹅，就可以证明“所有天鹅都是白色的”这个说法绝对是错误的。（至少，我们需要换一种更加严谨的说法：“欧洲所有的天鹅都是白色的。”）

同样道理，认真研究宇宙大爆炸理论，就会发现我们永远无法证明这是宇宙起源的精准模型。但是，我们有可能轻松地找到证据，证明这个理论是不正确的。自20世纪50年代以来，这样的证据就出现过好多次，宇宙学家只好结合这些新证据，修改并不断完善大爆炸理论。（天体物理学家弗雷德·霍伊尔觉得非常委屈，他认为自己的稳恒态理论从未得到同样的机会。此外，他还通过巧妙的证明向世人表明，他可以通过修正让这套理论与新数据高度吻合。）目前，大爆炸理论在几个问题上与数据的吻合程度非常高，但是，它随时可能被证明是错误的。也就是说，大爆炸理论经过了波普尔简化理论的检验。

对于那些通过层层叠加数学内容的方式构建而成、与现实之间联系不紧密的理论，情况就不一样了。按照波普尔简化理论的标准，弦论还不是（或许永远也不会是）得到普遍认可的科学理论。但是，这并不意味着弦论就没有研究价值。通过研究，人们有可能发现该理论的某些预言可以帮助我们完成证伪工作。尽管数百名科学家已经努力了几十年，但是以弦论目前的发展状态来看，人们仍然没有办法驳斥它。我们有理由认为弦论这个名称不是很合适，至多可以称之为“弦假设”。理论必须是可以检验的，而假设只是一种想法，不需要通过严谨的步骤来证明。

我们可以通过文字游戏，建立一个无法用波普尔简化理论来检验的科学理论。如果把上例中的天鹅理论改成“存在黑天鹅”，我们就没有办法证明它是错误的了。但是，如果我拿出一只黑天鹅，就可以证明这条理论是正确的。这个事实说明这条理论非常简单，只有非常简

单的概念才可以使用这种正话反说的办法。在这种情况下，我只需要检验某个命题是否有效，而且通过直接观察就可以达到这个目标。但是，进入粒子物理、宇宙学等领域之后，所有的证据都是间接证据。因为没有办法直接接触、做实验，所以我们没有办法证明某个事物真的存在。（例如，寻找希格斯玻色子的工作就具有这个特点。）在这种情况下，一个理论必须可以被证明有误这个标准具有非常重要的意义。

为了应对人们怀疑弦论割裂了数学与现实之间联系的喋喋不休的指摘，弦论研究者辩解称多余的维度都蜷缩到我们看不见的程度。但是，这个说法无法解决结果过多的问题。数学是不是终于发展到了过于偏离现实的地步？科学的目的是帮助我们理解、解释宇宙的运行规律，但科学是否已经忘记了它的“初心”呢？

第15章
数学的力量?

Are Numbers Real?

我们对数学与现实之间关系的探讨已经接近尾声，现在，我们不可避免地要提到一个现象。有人认为，至少物理学领域目前正处于本末倒置的状态——数学占据了主导地位，长此以往，很容易导致令人不安的后果，即科学家的研究将越来越难以被普通人理解。

物理学家尤金·维格纳讲过一个故事。两名昔日中学好友在一起聊天，其中一个人是统计学家，正在自豪地介绍自己的工作。他拿出一篇介绍人口变化规律的论文，然后告诉他的朋友，根据某种分布类型（高斯分布）可以预测人口将发生哪些变化。在这个过程中，他不可避免地要解释这篇论文里出现的大量晦涩难懂的数学符号。

朋友认为统计学家在论文里描绘的那幅图其实是一堆抽象数字的直观表示，对此他产生了怀疑：怎么能用这幅图来预测一群人，一群活生生的、有自己想法的个体的行为呢？但是，他发现这些数学符号中还隐藏着更加令人难以置信的东西。他指着论文中的一个符号，询问它的含义。统计学家说："这是π。你应该知道π的含义，就是圆的周长与直径之比。"朋友摇了摇头："原来你真的在捉弄我。人口怎么可能与圆的周长有关呢？"

维格纳是在一个讲座的开场部分讲述了这个故事的，目的是解释“数学在自然科学中不可思议的有效性”。他认为，这种不可思议的有效性表现在两个方面。第一，数学可以出人意料地应用于某些看似不相关的领域（例如，在研究人类行为时，π的出现就会令人感到奇怪）。第二，我们不能因为这些数学概念具有相同的规律，就断言这些数学概念与现实之间存在某种联系。这可能只是一种巧合，而我们在实验时正好碰上，因此效果不错。但是等到明天，或者当我们将它应用于另一种情况时，它也许就不适用了。

现在，我们可以回过头来看看数学的诞生过程。在本书前几章里我告诉大家，数字在刚开始的时候是用来表示实物的。最初，数字可能与计算山羊数量的手指相匹配，然后又与计算其他事物数量的手指相匹配，这可能是数字的第一个抽象化过程。接着，数字进一步抽象化，变成了表示手指的符号。然而，在这个阶段，数字与实物之间的关系仍然非常清楚、直接。随着数学的发展，负数使数学与现实之间的关系渐渐疏远（负数就相当于从整体中取走的物体的数量），随后数学领域又引入了虚数、$\aleph_0$等数字。维格纳说：

> 大多数更高层次的数学概念……被精心设计出来，这是因为这些数学概念是数学家展示自己的创造力和对形式美的品鉴能力的理想平台。

我们看到数学家正在不断突破可能性的限制，尽可能地拓展数学的应用领域，提出了一系列在逻辑上不会产生冲突的概念。即使搭建而成的完整结构在现实世界中没有实用价值，也找不到与之匹配的对象，他们也乐此不疲。与此同时，他们还为自己的所作所为感到震惊、困惑。之所以有这种感觉，是因为这些数学家（他们也是凡人）在创造一个个小世界，并取得数量众多而且和谐统一的成果（尽管有时候需要修改规则，例如将1从素数集中剔除）。与此同时，这个过程

也会让外行人感到困惑，因为他们认为有的成果除了可以用来炫耀自己的智商以外毫无意义，却仍然有人愿意耗费时间和精力，从事这方面的研究。

但是，纯粹以实用性来引领研究方向是不切实际的，科学研究如此，数学也如此。我们必须赋予数学家做实验的自由，因为我们不知道数学上含混不清的辩解之词何时会变成实用的工具。这个特点让那些政客以及负责为科学和数学拨付研究资金的其他局外人感到特别为难，他们觉得自己划拨的那些钱款似乎是供人“玩乐”的。如果拨款对象从事的是纯粹的数学研究，就等于为他们研究那些不切实际的抽象概念买单。但是，我们根本不知道这些概念什么时候会发挥作用。

数学家和科学家就像收藏家一样，他们如饥似渴地把那些看似无用的东西收藏起来，希望有一天这些东西会变成无价之宝。一旦如此，它们就会产生深远的影响。20世纪初，数学界以外的人几乎都不清楚非欧几何的发展前景——当然，爱因斯坦不在此列。除了可以在地球的弯曲表面上确定航行方向以外，非欧几何似乎与现实没有任何联系。但是，爱因斯坦创立的广义相对论却离不开它。

数学家在研究数学时无拘无束，根本不需要考虑其是否与现实有关。实用价值不应该成为评判数学研究的基本标准，就像我们不能根据载人飞船太空探索的副产品（为把人类送入太空所投入的智力和财力给社会带来的额外好处）的多少来评判这项活动一样。在这些副产品中，有的是实实在在的好处，但影响力通常比较小，有的则是不合逻辑的建议。人们常常罗列出GPS（全球定位系统）、气象卫星、空间望远镜等，作为支持载人航天活动的理由。但是，实际上，无须花费大量财力，以及执行各种危险的太空任务，人们也可以得到这些副产品。

不仅如此，根据副产品多少进行评判有时候还会得出彻头彻尾的错误结论。我曾经看过，有人认为广泛用于煎锅、水管、尼龙搭扣、

个人计算机等方面的不粘材料聚四氟乙烯要归功于美国国家航空航天局。事实上，前两项应用完全是普通研发工作的成果，只不过在几十年之后，美国国家航空航天局碰巧应用了这两项成果。美国国家航空航天局的确想减小计算机的体积，以便将它装进太空舱中，但是个人计算机发展的最大动力绝对不是美国国家航空航天局的需要，而是潜在的大规模市场。就效果而言，大规模市场的推动力远远大于小规模的专业应用。因此，我们必须忘掉这些副产品带来的“好处”，专心致志地考虑载人航天飞行的真正目的：它既是一项光荣的冒险活动，又有可能为人类带来生机。

与之类似，在评判数学家的抽象数学研究时，我们也可以列举出纯粹数学在应用方面的诸多“副产品”，但是大多数数学家从事相关研究却另有目的。过于关注实用价值，可能不利于取得重大突破。数学家希望迎接挑战，乐于享受在精神世界建功立业的乐趣。但是，物理学家通常会受到现实的羁绊，无法在梦想的国度里长时间地遨游。理论上，我们好像也可以创立一门专门研究纯理论的“异种物理学”（或许真的有这样的物理学，但是我电脑上的拼写检查程序却认为这个单词不存在）。事实上，物理学使用的很多模型都非常简单，与现实之间几乎没有多少相似之处，但是物理学研究的目的是预测并解释自然界的各种行为。在借山羊时，人们会用手指表示山羊的数量。同样，研究自然的物理学也不能脱离与自然的联系。

想要更好地理解物理学的研究内容，必须先了解模型的概念。只要有机会，我就会讲述一个在科学家群体中流传的古老笑话，从中一方面可以清楚地看出科学模型的本质，另一方面可以看到内行人和外行人对科学的不同理解。听到这个笑话后，外行人的反应最多是礼貌地一笑了之。笑话涉及的三个人分别是遗传学家、营养学家和物理学家，他们正在讨论赛马的培育方法。遗传学家说：“我们必须根据马的特性选择合适的马匹开展育种工作，多代之后才能培育出最优秀的赛马。”营养学家听后说道：“不对，更重要的是看马在成长过程中营养

是否均衡，能否增强肌肉。”物理学家听后摇摇头说：“我们可以假设赛马是一个球面。”

简单化的模型的确可能引人发笑，但其中并不是没有道理可言。上学期间，我一看到物理题目中出现“假设没有摩擦力和空气阻力”就感到十分生气，觉得他们在骗我，因为摩擦力和空气阻力肯定存在。既然可以这样假设，那么我们也可以“假设我知道正确答案”啊！然而，这也说明了一个事实：数学家在数学世界中拥有至高无上的控制权，而物理学家必须更注重实际。我们以身体为例。物理学在研究这个我们都非常熟悉的对象时，必须使它简单化，而不是研究一个个原子。原子是量子，具有概率特性，而且原子的数量非常庞大，我们身体里大约有10^{27}个原子。任何人都不可能合理地预测每个原子的行为，因此我们只能将所有原子视为一个整体，并在此基础上建立模型。

鉴于我们周围的世界无比繁复，物理学竟然可以做出各种预测，这实在是一件匪夷所思的事。但是物理学真的做到了，其主要手段就是建立简单化的数学模型。在理论上，这些模型不一定真的有用。但是，我们发现宇宙万物之间具有某种协调性，从而为物理学家提供了某种帮助。如果被我们视为基本常数的值（例如光速）不断变化，物理学就会遭受灭顶之灾。所有的物理系统都会受到灾难性的打击，我们甚至也无法预测任何事物的发展趋势，因为事物发展将表现出一种茫然无序的情况。

我们通常认为，地球上的规律同样适用于其他地方，但现实似乎并非如此。比如，我们知道物体在月球上比在地球上轻得多。因此，机器在月球上的性能也会大不相同。但是，牛顿的一个伟大创新就是大胆地假设引力是一个普遍现象，他认为月球上的引力作用与地球上的一模一样。截至目前，除了几个可能的例外情况，这个假设似乎并

无不妥之处。我们无法证明这是一个事实，但是，如果我们不做出这样的假设，科学研究的所有努力都将毫无意义。

值得庆幸的是，很长时间以来，这种假设经受住了考验，这就意味着我们可以利用数学模型来研究现实，并且有可能取得令人吃惊的成功。在这个过程中，我们常使用不变性原理（简而言之，不变性原理认为我们建立的物理“定律”适用于所有时空）。

尽管科学的不变性不受实验操作者的影响（实验者的性别、年龄或种族不会影响实验结果），但并不是所有“实验”在完成质量上都是相同的，因为实验者采取的控制措施各不相同。然而，其中的道理并不十分明显。那些热衷于顺势疗法，或者认为自己对电磁辐射敏感，或者相信自己能看见鬼魂的人有时迷惑不解，原因就在于此。科研人员通常会花大力气将某个现象分离出来，以便从这个现象中提取数据时不会受到外界因素的影响。这在实验室环境中是可以做到的，但是一旦离开实验室，效果就会大打折扣。

比如，2014—2015年第二代宇宙泛星系偏振背景成像微波望远镜的发现，一度被视为引力波，但是后来人们发现那些数据其实受到了银河系中尘埃的影响。任何天文学和宇宙学观测都有可能受到意料之外的干扰，我们的日常体验亦如此。科学界有一句老话：“传闻再多，也不是数据。”即使我们有某种亲身经历，也不意味着大脑对这些经历的解读就是有效的科学模型。即使我采取顺势疗法之后感觉很好，我也无法判定自己是在这个疗法的帮助下恢复健康的，因为我没有合适的比较标准。我感觉不错的原因可能有很多种。最有可能的原因是我的健康状况真的有所好转，也有可能只是我的感觉很好，但身体状况并没有真正得到改善。也就是说，我们必须清醒地认识到，不变性假设并不意味着我们可以把道听途说当作科学证据。

尤金·维格纳指出：“自然律只在特殊条件下，即世界现状的所有相关决定因素均为已知时，才能用于预测未来。”一般而言，这个条件

永远无法实现。我们无从了解我们所面对的所有条件，因此我们只能依赖假设和简单化处理。这样一来，我们使用的模型与现实越来越脱节，得出不正确结果的可能性也越来越高。

即便如此，我们仍然可以看到，在某些领域，研究人员没有依靠与现实相匹配的模型，而是借助抽象数学做出了与现实吻合的预测，从而证明数学是一个效果惊人的工具。维格纳以兰姆移位（原子中两个电子能级之间的微小差别，这一发现推动了量子电动力学的发展）为例，说明了这个问题："由贝特提出、施温格创建的兰姆移位量子理论是一个纯粹的数学理论，它唯一的直接贡献就是证明存在某种可测量的效应。事实表明，实验结果与理论计算的吻合程度超过千分之一。"

尽管这些方法取得了成功，但这并不能保证这些模型一定可以发挥作用。建立模型时，我们总会依赖自己的经验。比如，在研究物理学的两大支柱——始于20世纪的量子论和相对论时，人们依赖于两种不同的数学结构，但是这两种结构无法融合。几乎所有的物理学家都认为肯定有办法让它们融为一体——要么修改其中一个理论，使其适合另一个理论的数学结构，要么设计出同时适用于两大理论的新结构。但是，现有模型很有可能就是效果最好的可行模型，而且它们永远无法实现融合，"万能理论"也永远不会出现。

我们并不一定非得建立一个可身兼数职的普适型数学系统。尽管有人会提出相反的观点，但我仍然认为宇宙的本质不是数学。只不过，某些（并不是大多数）数学内容的基础是现实世界的观察结果。实际上，数学是一个功能强大的工具，可以帮助我们建立宇宙模型，而模型一定有其局限性。在这里，我要再次引用维格纳的话："我们根本不知道我们的理论为什么如此有效，因此，它们的精确性并不能证明它们的真实性与一致性。"维格纳有力地指出，即使基于数学的理论可以做出有效的预测，也不一定意味着它具有某种价值。

这个观点似乎非常奇怪。既然数学可以有效地预测未来，为什么否认它是现实世界的精确表达呢？原因在于我们已经发现的一个事实：几乎所有的物理学内容都离不开大量的简单化处理与假设。在我们建立模型时，宇宙黑箱里真实存在的很多元素都有可能被我们忽略。举一个非常简单的例子，下面是一个预测明天早晨太阳是否会升起的计算机程序：

```
如果年份<3000
输出“太阳将会升起”
否则
输出“太阳不会升起”
结束
```

这其实是一个数学模型，但大多数人更熟悉计算机程序的逻辑结构，而不熟悉数学的符号结构。我可以告诉大家，在公元3000年之前，这个模型的预测结果都不会有任何错误，但是到了公元3000年，它就会出错（我希望如此）。你也许会说，我输入的具体日期与现实之间没有任何联系。但是，从一定意义上说，这就是我举这个例子的目的所在。我们也不知道自己输入科学模型中的那些常数、公式是否与现实有关联，我们只知道其预测结果与我们观察到的现实非常吻合。但是，这些模型的内容与现实之间的联系，不一定比我编写的那个高度精确但是毫无价值的计算机程序与现实之间的联系更紧密。

本书谈到了很多了不起的数学家所提出的崭新的、有影响力的数学概念，其中一些概念是作为纯粹数学的推论被提出来的。例如，16世纪，吉罗拉莫·卡尔达诺提出了虚数的设想，当时他并没有考虑虚数会有什么实用价值。但是，在科学家和工程师的手里，虚数变成了一个功能强大的工具。数学家抱着开发一个新的实用方法的目的，有意

识地开发出一种数学方法的情况并不多见，其中最显而易见的例子就是微积分的前身——牛顿的流数术。

为了揭示数学的本质，了解数字是否真实存在，我们可以把数学家视为古时候的铁匠。铁匠的任务是为社会提供工具，为其他行业制造必需的设备。但是，他们也是艺术家，有时还会别出心裁，用金属打造一些用途不太明确的物件。与铁匠相似，数学家也为科学研究提供了大量工具，尽管他们可能对由纯粹数学构成的抽象世界更感兴趣。

借助数学，科学已经取得了一系列成功，但是我们有理由相信人们在考虑数学的应用性时步子可能迈得太大了。美国物理学家理查德·费曼曾经把那些伪科学家徒有其表的科研活动称作“草包族科学”。但我认为用这个词表达另外一种意义，效果可能更好。据称，20世纪上半叶，美拉尼西亚群岛上的居民由于货机崇拜而将飞机模型与真正的飞机混为一谈（对于货机崇拜这个现象，历史学没有确定的结论，反而是神话传说表现出一副言之凿凿的样子）。我认为，有些科学家同样把他们的模型与现实混为一谈了。

数学一度被视为帮助我们了解、解释物质世界的工具，但它现在已经成了一个独立的实体。在用数学推演出结果之后，我们会绞尽脑汁地让它们与观察结果一致。例如，数以百计天资聪颖的人正在研究弦论，但是他们的努力很有可能会徒劳无功。物理学教授萨宾·霍森菲尔德称：“在这个过程中，很多物理学家开始相信，仅凭纯粹的（数学）逻辑和一定的美感，就有可能在没有观察结果的前提下提出一条理论。他们肯定认为他们的大脑和宇宙之间存在某种神秘的联系，仅凭思考就可以发现自然律。”

当然，物理学不是科学的全部，但是在运用数学工具这个方面，其他学科仍然不及物理学。我们已经知道，那些从事心理学等软科学研究的人往往不能得心应手地使用统计学等工具。而且，由于这些工

具看似简单，但是常常与直觉相悖，因此他们经常会得出一些错误的结果。有些科学领域还需要进一步加强对数学的了解，以便为观察结果建立合适的模型。但是对于物理学而言，花在“集邮”上的时间或许应该多一点儿，而躲在象牙塔里钻研数学的时间或许应该少一点儿。

智能手机等日常设备应用了量子技术，其背后理论所涉及的数学知识是99%的人都无法理解的。但是，我们同样需要通过加强对实验的倚重、减弱对数学理论的依赖，建立一些不同的、更容易解释的模型。归根结底，现实世界的科学模型并不全是数学层面的，有的模型还需要反映我们对周围世界的观察结果，以便帮助我们更好地理解现实。建立模型绝对不止一种方法。

统计学家早就警告我们，相关性不等于因果关系。两个数据集在一段时间里同时发生变化，并不代表它们之间存在因果关系，也不代表未来它们会继续保持这种相关性（要确保相关性持续存在，两者之间就必须存在因果关系）。“二战”之后，英国香蕉进口贸易与妇女怀孕就是有相关性但是没有因果关系的一个经典例子。当时，香蕉进口量增加，英国怀孕妇女的数量也会增加，反之亦然。但是，任何头脑清醒的人都不会认为妇女怀孕的原因是香蕉进口贸易（事实上，这两件事可能都与第三个因素相关）。一般而言，我们遇到的相关性都远没有这个案例那么荒诞不经，因此我们更有可能受到表象的迷惑，误以为存在因果关系。

不存在因果关系的相关性案例不胜枚举，有一个网站甚至还将这些相关性案例列举出来。这个网站告诉我们，在10年时间里，美国上吊自杀的人数与美国政府拨付给科学研究、空间探索和技术发展等方面的经费之间存在显著的相关性；同一时期，美国缅因州的人造奶油人均消费量与离婚率之间存在显著的相关性。事实上，这些相关性显然毫无意义。然而，如果新闻节目主持人告诉我们政府的新政策导致

股市下跌，我们可能根本不会感到惊讶，尽管这种因果关系也只是一个假设而已。

科学家喜欢在实验室条件下做实验，原因之一就是他们在实验室里可以控制很多在“自然环境中”无法控制的可能变量。在实验室中，因果关系比较容易确定，但是在现实中，发现因果关系的难度要大得多，因为诸多干扰因素会把某个现象的真正原因隐藏起来。

正因为因果关系难以确定，所以人们很难判断不同的饮食方法到底会对人类健康产生什么影响。例如，科学家可能会注意到大量食用西红柿的人患心脏病的概率低于普通人，但是我们不能就此认为只要多吃西红柿就一定能改善我们的健康状况。我们长期生活在一个异常复杂的世界里，这与控制措施严密的实验室不同。我们将发现，大量食用西红柿的人与那些常吃垃圾食品的人还存在很多其他不同点，其中最重要的不同点或许与西红柿根本没有关系。

如果我们可以对几千人做实验，控制他们的饮食，那么几个月之后，我们或许可以进行严谨的科学分析。但是，事实上，大多数饮食结构研究都需要综合考虑一系列的差异，操作起来难度很大（往往还需要依赖实验对象极不准确的自我报告），因此很难保证实验结果的精确性。

在实验室里工作的物理学家无须面临如此恶劣的条件，但是在难以实施控制的科学领域（例如宇宙学），研究人员却会遭遇同样的问题。在实验室中，科研人员也可能遇到异常复杂的情况，或者需要通过非常曲折的方式进行间接观察。例如，在大型强子对撞机项目中，探测器提供的探测结果就非常复杂、极其混乱。在这种情况下，我们很有可能受到诱惑，忍不住使用某种数学方法，原因是这个方法“似乎是正确的”，它给人一种难以抗拒的美感，而不是因为观察结果要求我们必须采用这种数学方法。结果，虽然我们的数学模型可以产生大量

与实验结果相吻合的数据，但是模型本身却与现实没有任何联系。对数学的过分依赖，已经引起了若干当代物理学家的关注。

在科学史上，我们可以找到大量模型与现实脱节的实例，例如，托勒密天文学中使用的本轮系统。这套系统的基本原理是，根据观察结果建立相匹配的数学模型，并对其进行完善。这套系统的应用时间为1 300多年，而且由于不断完善，它与观察结果的吻合度一直非常高。但是，使用这种圆周旋转模型来描述行星的运行方式的做法，并没有充分的科学理由。这个模型基于一个错误的物理假设（地球是宇宙的中心），因此它无法逃脱覆灭的命运。毕竟，即使数学工具给出的答案与观察结果高度吻合也无济于事，因为仅仅得到数学的支持是不够的，数学不可能拥有决定一切的权力。

数学鸠占鹊巢，把实验挤到了次要位置，这让在著名的圆周理论物理研究所担任主任一职的尼尔·图罗克难以接受。图罗克曾发表了下面这番言论：

> 自然为我们提供了这些不可思议的线索，但是我们并没有理解其中的含义。事实上，我们正在做着南辕北辙的努力，导致我们的理论越来越复杂、越来越不自然。我们搬来了更多的场、维度、对称，想尽一切办法解决这个问题，却没有解释最基本的事实。

实际上，图罗克批评的是人们利用数学来推动科学研究这个现象。从本质上看，数学研究的是真理和事实，这是数学与科学的区别之一。在数学系统里，事实不容置疑。例如，在传统算术中，2加2一定等于4。这是算术法则规定的事实，是由这套数学系统的本质决定的，任何人都不可能找到证据驳斥它。科学家兼科幻作家艾萨克·阿西莫夫说：“随着时间的推移，人类活动的所有领域几乎都会发生显著的变化，这些变化可以被视为修正或者拓展，又或者兼而有之……现

在，我们可以看到数学到底有什么独特的地方了。数学这个领域从来没有进行过重大的修正，所有的改变都是拓展。”

在数学领域，我们必须注意语言的精确性。如果我们换用另一个数基（比如基数为3时，2加2等于11），或者使用第1章里讨论的算术法则（数字不是一直往上增加，而是像钟面一样，达到最大值之后就会从头再来），那么2加2完全有可能得出不同的结果。对于数学家而言，传统算术与钟表算术不存在谁更“真实”的问题，尽管一种算术适用于所有传统实物，而另一种则只能用于处理周期性事件。在传统算术的特定系统中，我们绝不能背离在该系统中发挥基础作用的所有事实。

然而，科学与数学在这方面有所不同。当科学家讨论137亿年前使宇宙开始膨胀的那次大爆炸时，他们描述的内容与我们通过书写2 + 2 = 4这个等式表达的内容有所不同，因为后者是一种事实，而宇宙大爆炸则是根据当前数据建立起来的最可靠的理论。在我创作本书期间，当得起“最可靠”这项殊荣的大爆炸理论至少被修改过三次，每次都是因为有数据证明当前版本是错误的。在未来的某一天，大爆炸理论甚至有可能被全盘否定，并被一个更可靠的理论取代。

科学一直都是临时性的。科学的目的不是寻找绝对真理，科学的核心要素也不是事实，但这并不是说事实在科学中不重要。科学研究中涉及大量的事实收集任务，也就是遭到卢瑟福诟病的“集邮”工作，部分原因在于很多工作不过是贴上“已成事实”这个标签。很多事实都属于可观测事实，包括我在打字时会使用一个键盘、我的计算机需要电源等事实。当科学提供解释性理论时，我们必须清楚，此时事实已经离我们远去了。比如，当我说光是一种波、一束粒子流，或者量子场的波动时，我描述的其实不是光，而是一个数学理论，或者是在进行类比。我们可以说光或者原子是存在的，也可以创建理论描述它们

的本质或者作用原理（因为它们的规模及作用环境与我们可以观察的宏观世界大不相同），但这两者并不是一回事。

科学家经常忽略甚至忘记科学与真理之间的间接关系，或许是因为这种关系会造成某种危险，让人们以为所有的想法和理论在重要性方面都不相上下。因为我是一名科普作家，所以经常有人向我介绍各种各样的理论，甚至宣称他们已经证明爱因斯坦的相对论是不正确的。还有一些人根本不相信科学，而是对巫术深信不疑，坚信能量可以无中生有，坚信顺势疗法可以治病。但是，在面对多种可能性时，科学没有采取一视同仁的态度。所有人都可以对爱因斯坦的观点提出质疑，但是就目前而言，相对论取得了非常好的效果，除非有新的令人信服的证据可以证明某个理论与实验结果或宇宙观测结果的契合程度更高，否则人们不会转而接受其他理论。要彻底推翻科学，用巫术取而代之，需要更有说服力的数据。

有人认为科学就是数学，或者数学就是宇宙的本质。尽管其中不乏才高八斗、智力超群之人，但是我认为，从数学与科学本质上的不同可以看出他们都错了。这两个学科，一个学科（数学）包含一系列事实，我们可以根据我们制定的法则确定它们都是事实；另一个学科（科学）则涉及一系列模型和理论，我们可以利用数据测试这些模型和理论，但是绝不可以称之为事实真理。

在确定数学这门与现实世界脱节的学科应该被赋予的地位时，我们需要考虑一个非常重要的因素。在脱离现实的那个绚丽的世界中，数学家可以编织神奇的魔术，以数学为基础创造出所有事物，但这并不是科学研究应用数学工具的真正目的。我们在第2章提过的数学家理查德·哈明说：

> 使用数学工具时，我们都会有所选择。数学工具不是通用的。在发现标量不适合表示作用力之后，我们就发明了矢量这个新的数

学工具。接着，又发明了张量……我们根据具体情况选择不同的数学工具，因为同一个数学工具不可能适用于所有情况。

科学已经取得了辉煌的成就，而且未来的前景更加光明。数学也已被证明是一个功能强大的工具，可以帮助我们建立宇宙模型，还将继续扮演强大工具的角色。在运用得当的情况下，数学仍然是帮助我们理解物理学基本原理、探索宇宙奥秘的最有效方法，但这并不意味着我们可以滥用这个工具。科学界还需要注意另外一个同样重要的问题：不可将模型与现实混为一谈，而要时刻牢记数学世界别有洞天，有时候它像一面镜子，能把现实世界的情况准确地反映出来，但这并不意味着它就是现实。像爱丽丝那样透过镜子就可以造访的数学镜像世界是不存在的。

某些数字和数学过程毫无疑问是真实的，至少与现实世界的真实物体和行为存在一一对应关系。自然数，即非负整数，肯定源自物理对象，并且与真实物体一样遵循相同的算术法则。人类花费了很长一段时间，才为负整数奠定了同样的理论基础，我们可以在计算电荷的过程中看到它们的身影。随着数学的进一步发展，数学与现实之间的差别越来越明显。例如，虽然分数和几何在现实世界中可以找到与之匹配的对象，但是它们高度精确，这与现实世界的混乱状况迥然不同。我们栖身的这个世界就像柏拉图的山洞一样，不可能找到线条宽度为0的几何图形，而且，由于现实世界是由原子构成的，我们也无法完全等分蛋糕。

诚然，有时候我们可以把某些东西分成精准的几等分，例如钱，25美分就正好是1美元的1/4。但是，这是因为钱从本质上看是量化的，有现成的分割方法，因此我们可以使用分数。然而，这种办法需要付出数学成本。我们可以精准地得到1美元的1/4，但是我们没有办法精确地得到1美元的1/3。尽管其数学表达非常简单，但是现实中根本不存在这个概念。

本书从数学专业人士的角度告诉大家，分数和几何学与美轮美奂、光怪陆离的数学世界相比，不过是冰山一角。数学世界无比广袤，即使数学家穷尽毕生精力，也很可能不会有任何真正的发现。但是，有的数学结构与数学机制的确与现实有相似之处。这些数字和程序也许并非真实存在，但是它们仍然可以帮助我们找到问题的答案。

尽管数学可以脱离现实，但是我们必须让实用数学建立在物质世界的基础之上，使科学可以被所有人接受。那么，数字到底是不是真实存在的？我认为，数字从最基本的意义上看确实是真实存在的，但是大多数的数学内容则与之相反。数学就是一个梦幻世界，有时与我们身边的这个世界非常相似，是这个世界的完美反映，可以作为我们理解现实世界的工具。但是，所有的数学工具都必须恰当运用。只要我们（以及科研人员）牢记这一点，就不会犯下大错。

致谢

感谢圣马丁出版社每个为本书的出版做出努力的人，这其中包括迈克尔·霍姆勒、劳伦·雅布隆斯基。同时也感谢每个帮助我思考数学与科学之间关系的人，包括牛津大学知名教授阿德里安·穆尔。他之所以研究哲学，我之所以对数学和无穷大产生兴趣，都是因为受到了同一个人的启发——曼彻斯特文理学校的尼尔·谢尔登。

注释

第1章 虚拟的“居民”？ ㊟

4—The article on poor attitude to math in the UK is Wendy Jones, “Bad Attitudes to Maths Makes Children Switch Off,” *Guardian,* March 7, 2012, accessed July 12, 2015, www.theguardian.com/teacher-network/teacher-blog/2012/mar/07/world-maths-day-adult-innumeracy.

5—St. Augustine’s damnation of mathematicians is from St. Augustine, *Genesi ad Litteram* (*The Literal meaning of Genesis*), trans. Roland J. Teske (Washington: CUA Press, 2010), p. 76.

6—The British Court of Appeal discussion of the meaning of the number one is reported in Steve Connor, “What Exactly Does ‘One’ Mean? Court of Appeal Passes Judgement on Thorny mathematical Issue,” *Independent*, accessed June 30, 2015, www.independent.co.uk/news/science/what-exactly-does-one-mean-court-of-appeal-passes-judgement-on-thorny-mathematical-issue-10350568.html.

9—Steven Weinberg's comparison of the International Space Station and the Superconducting Super Collider was made in a telephone interview with the author in February 2013.

10—Max Tegmark's assertion that the universe is made of mathematics is from Max Tegmark, *Our Mathematical Universe* (London: Penguin Books, 2015), pp. 243–318.

第2章 史前人类的计数系统

15—The Ishango bone is described in Amir Aczel, *Finding Zero* (New York: Palgrave, 2015), pp. 20–21.

17—Details of the variants of Uruk number systems from Leonard Mlodinow, *The Upright Thinkers* (London: Allen Lane, 2015), p. 49.

18—Richard Hamming's amazement at the abstraction of integers is described in Richard Hamming, "The Unreasonable Effectiveness of Mathematics," *American Mathematical Monthly* 87, no. 2 (February 1980), pp. 81–90.

22—Information on Ancient Greek numbering from Amir Aczel, *Finding Zero* (New York: Palgrave, 2015), p. 14.

第3章 毕达哥拉斯：万物皆是数字

30—Theories for the reason the Pythagoreans avoided beans are discussed in Charles Seife, *Zero: The Biography of a Dangerous Idea* (London: Souvenir Press, 2000), p. 25.
30—Details of the meanings ascribed to numbers by the Pythagoreans from Brian Clegg, *A Brief History of Infinity* (London: Constable & Robinson), pp. 34–35.

31—The prediction of a tenth heavenly body by Philolaus is described in Steven Weinberg, *To Explain the World* (London: Allen Lane, 2015), p. 78.
37—Details of Zeno's paradoxes from Brian Clegg, *A Brief History of Infinity* (London: Constable & Robinson, 2003), pp. 10–17.
38—The Greek consideration of numbers as collections of objects and fractions as parts is described in David Fowler, *The Mathematics of Plato's Academy* (Oxford: Oxford University Press, 1999).

第4章 欧几里得：几何定理的完美证明

47—Information on Euclid's work from *Euclid's Elements: All Thirteen Books Complete in One Volume,* trans. Thomas Heath (Santa Fe, NM: Green Lion Press, 2002).
52—Details of the Ancient Greeks who spent their time attempting to square the circle from Brian Clegg, *A Brief History of Infinity* (London: Constable & Robinson, 2003), p. 68.

第5章 阿基米德：用沙粒填满宇宙

56—The description by Plutarch of Archimedes considering mechanical devices to be geometry at play is quoted in Thomas Heath, *The Works of Archimedes* (New York: Dover, 2002), p. xvi.
57—The suggestion that Archimedes was related to the royal family of Syracuse is from Thomas Heath, *The Works of Archimedes* (New York: Dover, 2002), p. xv.
57—The quote from *The Sand-reckoner* on sand location is from Thomas Heath, *The Works of Archimedes* (New York: Dover, 2002), p. 221.

58—Archimedes quote from Aristarchus on the size of the sphere of the fixed stars is taken from Thomas Heath, *The Works of Archimedes* (New York: Dover, 2002), p. 222.

第6章 斐波那奇：阿拉伯数字的登场

65—The views of Brahmagupta and Bhāskara on zero divided by zero are from Brian Clegg, *A Brief History of Infinity* (London: Constable & Robinson, 2003), p. 111.
66—The introduction of the BC/AD dating system is discussed in Bonnie Blackburn and Leofranc Holford-Strevens, *The Oxford Companion to the Year* (Oxford: Oxford University Press, 2003), pp. 767–82.
68—The Old Babylonian tablet with a value for the square root of 2 is described in Carl B. Boyer and Uta C. Merzbach, *A History of Mathematics* (New York: John Wiley & Sons, 1991), p. 27.
71—The mention of Hindu numerals by a Syrian bishop in 662 is noted in David Eugene Smith, *History of Mathematics,* vol. 1 (New York: Dover, 1958), p. 167.
72—The translation of the Cambodian inscription with a zero dating to AD 683 is in Amir Aczel, *Finding Zero* (New York: Palgrave, 2015), pp. 94–97.
73—The story of Amir Aczel's successful attempt to find the oldest known zero is in Amir Aczel, *Finding Zero* (New York: Palgrave, 2015), pp. 98–178.
73—The original Indian use of the same symbol and concept for a placeholder zero and for an unknown quantity is described in Robert Kaplan, *The Nothing That Is* (London: Penguin, 2000), pp. 57–59.
74—The different interpretations of zero from classical Indian mathematicians are taken from Robert Kaplan, *The Nothing That Is* (London: Penguin, 2000), pp. 72–73.

75—John Donne's railing against "nothing" is from John Donne, *The Works,* vol. 6 (London: Parker, 1839), p. 155.

75—The resistance to use of numerals rather than words in Florence and Belgium is described in Robert Kaplan, *The Nothing That Is* (London: Penguin, 2000), p. 102.

77—The Ancient Greek style word form of the equation A+B=C+D and their approach to numerals is described in Reviel Netz, *The Shaping of Deduction in Greek Mathematics* (Cambridge: Cambridge University Press, 1999), p. 98.

77—The modern rendering of the approach taken by Diophantus to an algebraic equation is from Carl B. Boyer and Uta C. Merzbach, *A History of Mathematics* (New York: John Wiley & Sons, 1991), p. 181.

78—The suggestion that algebra may have been developed to help with complex inheritance rules is in Carl B. Boyer and Uta C. Merzbach, *A History of Mathematics* (New York: John Wiley & Sons, 1991), p. 233.

第7章 培根：数学是自然科学的钥匙

79—The quote from Roger Bacon on the importance of mathematics is from Roger Bacon, *Opus Majus,* trans. Robert Belle Burke (Kila, MT: Kessinger Publishing, 1998), p. 116.

88—Information on Muybridge's Zoopraxographical Hall at the World's Columbian Exposition from Brian Clegg, *The Man Who Stopped Time* (Washington, DC: Joseph Henry Press, 2007), pp. 225–26.

88—The book that temporarily ruined Muybridge's reputation as the grandfather of moving pictures was *A Million and One Nights* by Terry Ramsaye, discussed in Brian Clegg, *The Man Who Stopped Time* (Washington, DC: Joseph Henry Press, 2007), pp. 245–47.

89—Roger Bacon's claim that squaring the circle was "clearly understood" comes from Roger Bacon, *Opus Majus,* trans. Robert Belle Burke (Kila, MT: Kessinger Publishing, 1998), p. 15.

90—The quote on Bacon's support of math from John Wallis is from Joseph Frederick Scott, *The Mathematical Work of John Wallis* (New York: Chelsea Publishing Company, 1981), p. 142.

90—Roger Bacon's extended eulogy to the importance of mathematics is from Roger Bacon, *Opus Majus,* trans. Robert Belle Burke (Kila, MT: Kessinger Publishing, 1998), pp. 116–17.

92—Information on Bradwardine and Oresme from Carl B. Boyer and Uta C. Merzbach, *A History of Mathematics* (New York: John Wiley & Sons, 1991), pp. 262–66.

第8章 高斯：神通广大的虚数

98—Information on Cardano and Tartaglia from Carl B. Boyer and Uta C. Merzbach, *A History of Mathematics* (New York: John Wiley & Sons, 1991), pp. 282–89.
98—Cardano's remark that an imaginary number was "as subtle as it is useless" is quoted in Roger Cooke, *The History of Mathematics: A Brief Course* (London: Wiley, 1997), p. 310.

第9章 牛顿：微积分与宇宙观

102—The Marquis of Laplace wrote about an intelligence that could perfectly predict the future in Pierre Simon Laplace, *A Philosophical Essay on Probabilities,* trans. F. W. Trusscott and F. L. Emory (New York: Cosimo Inc., 2007), p. 4.

113—The assertion that Descartes primarily intended his analytical geometry as a way of constructing geometric forms, not of deducing algebraic forms is from Carl B. Boyer and Uta C. Merzbach, *A History of Mathematics* (New York: John Wiley & Sons, 1991), p. 337.

105—Details of the contents of Newton's library are from John Harrison, *The Library of Isaac Newton* (Cambridge: Cambridge University Press, 2008), p. 59.

110—Newton's letter to Leibniz with his coded claim is from Isaac Newton, *The Correspondence of Isaac Newton*, vol. 2, ed. H. Turnbull (Cambridge: Cambridge University Press, 1959), p. 134.

112—The quote from Berkeley on the errors in calculus is taken from Carl B. Boyer and Uta C. Merzbach, *A History of Mathematics* (New York: John Wiley & Sons, 1991), p. 430.

第10章 卡尔达诺：概率与“水晶球”

132—Maxwell's development of a distribution for the speed of gas molecules is described in Rhodri Evans and Brian Clegg, *Ten Physicists Who Transformed Our Understanding of Reality* (London: Robinson, 2015), p. 93.
135—The study showing that a double SIDS case would be expected in England every eighteen months was Stephen Watkins, "Conviction by Mathematical Error?," *BMJ*, 320, no. 7226 (2000), pp. 2–3.
136—Information on Isaac Asimov's use of mathematical prediction in the Foundation series from Brian Clegg, *Ten Billion Tomorrows* (New York: St. Martin's Press, 2015), p. 272.
138—The details of Linzmeyer's fifteen-card correct run and the quote from Rhine on the subject are from Joseph Banks Rhine, *Extra-Sensory Perception* (Hong Kong: Forgotten Books, 2008), p. 86.

第11章 麦克斯韦：关于电磁波的数学方程组

146—Biographical details on Maxwell and details of his electromagnetic theory from Rhodri Evans and Brian Clegg, *Ten Physicists Who Transformed Our Understanding of Reality* (London: Constable, 2015), pp. 252–71.

156—Frank Wilczek's assertion that the collection of fields that fill the universe form an ether is made in Frank Wilczek, *The Lightness of Being* (Philadelphia: Basic Books, 2008).

157—Details of retarded and advanced waves from Brian Clegg, *How to Build a Time Machine* (New York: St. Martin's Press, 2011), pp. 152–55.

第12章 康托尔：让一众科学家挠头的无穷大

161—Ancient Greek ideas on infinity are from Brian Clegg, *A Brief History of Infinity* (London: Constable & Robinson, 2003), pp. 29–32.
163—Gauss's dismissal of the reality of infinity was made in a letter to the Danish astronomer Heinrich Christian Shumacher, dated July 12, 1831.
163—The details of Galileo's work on infinity and his attempts to get the book containing it published are from Brian Clegg, *A Brief History of Infinity* (London: Constable & Robinson, 2003), pp. 80–92.
164—Quotes from Galileo's dedication are from the translation by Henry Crew and Alfonso de Salvio of Galileo Galilei, *Dialogues Concerning Two New Sciences* (New York: Dover, 1954), pp. xvii–xviii.
166—Sagredo's assertion about ants carrying a ship is from Galileo Galilei, *Dialogues Concerning Two New Sciences* (New York: Dover, 1954), p. 20.

172—The suggestion that ordinal numbers may predate cardinal numbers is from Carl B. Boyer and Uta C. Merzbach, *A History of Mathematics* (New York: John Wiley & Sons, 1991), p. 5.
174—The idea that basing the natural numbers on set theory abstracts it from reference to the real world is from Roger Penrose, *The Road to Reality* (London: Vintage, 2005), p. 65.
184—Gödel is quoted on the problems of the ZFC axioms in Natalie Wolchover, "To Settle Infinity Dispute, a New Law of Logic," *Quanta Magazine,* November 26, 2013, accessed June 30, 2015, www.quantamagazine.org/20131126-to-settle-infinity-question-a-new-law-of-logic/.
192—The quote from Max Tegmark on infinity is from Max Tegmark, "Infinity Is a Beautiful Concept—And It's Ruining Physics," *Crux,* February 20, 2015, accessed June 30, 2015, blogs.discovermagazine.com/crux/2015/02/20/infinity-ruining-physics/#.VZKZy2DL_gD.

第13章 爱因斯坦：量子物理与抽象数学

202—Einstein's description of having his 'happiest thought' while sitting in the Bern patent office is from his 1922 Kyoto Lecture and referenced in W. F. Bynum & Roy Porter (Eds.) *Oxford Dictionary of Scientific Quotations* (Oxford: Oxford University Press, 2005), p. 198.
207—Proposition 32 of Euclid's Elements is taken from *Euclid's Elements: All Thirteen Books Complete in One Volume,* trans. Thomas Heath (Santa Fe, NM: Green Lion Press, 2002), p. 25.

第14章 诺特：对称之美与隐形恶龙

225—Albrecht and Magueijo's paper on the possibility of a changing speed of light removing the need for cosmic inflation is Andreas Albrecht and João Magueijo, "Time Varying Speed of Light as a Solution to Cosmological Puzzles," *Physical Review D*, 59, no. 4 (1999), 43516, journals.aps.org/prd/pdf/10.1103/PhysRevD.59.043516.

226—Details of the Oklo natural nuclear reactors and their implications for symmetry of electrical charge through time from Leon Lederman and Christopher Hill, *Symmetry and the Beautiful Universe* (New York: Prometheus Books, 2004), pp. 40–43.

233—Leon Lederman points out the dangers of assuming perfect symmetries in Leon Lederman and Christopher Hill, *Symmetry and the Beautiful Universe* (New York: Prometheus Books, 2004), p. 19.

237—Martin Bojowald's quip that string theory is a theory of everything because everything and anything can happen is from Martin Bojowald, *Once Before Time: A Whole Story of the Universe* (New York: Alfred A. Knopf, 2010), p. 83.

237—The comment that string and M-theorists could be "away in Never-Never Land" is from Paul Davies, *The Goldilocks Enigma* (London: Penguin Books, 2007), p. 47.

238—The concerns about Popper's philosophy of science are taken from Tim Lewens, *The Meaning of Science* (London: Pelican Books, 2015), pp. 22–44.

第15章 数学的力量?

243—The story of the statistician whose friend doubts the validity of pi, and other thoughts on the nature of mathematics by Wigner are

from Eugene Wigner, "The unreasonable effectiveness of mathematics in the natural sciences. Richard courant lecture in mathematical sciences delivered at New York University," *Communications on Pure and Applied Mathematics* 13, no. 1 (May 11, 1959), pp. 1–14.
254—Sabine Hossenfelder's dismissal of deriving physical theory from aesthetics is from Sabine Hossenfelder (2015). "Does the Scientific Method need revision," accessed June 17, 2015, backreaction.blogspot .co.uk/2015/01/does-scientific-method-need-revision.html.
255—The apparent causal link between import of bananas and pregnancies is discussed in Brian Clegg, *Dice World* (London: Icon, 2013), p. 33.
255—The website providing apparent correlations between different sets of data with no causal link is Tylervigen.com, "15 Insane Things That Correlate With Each Other," accessed June 22, 2015, www .tylervigen.com/spurious-correlations.
256—The book giving details of scientific investigations of psi phenomena is Brian Clegg, *Extra Sensory* (New York: St. Martin's Press, 2013).
256—Details of how dietary science advice has problems in differentiating between correlation and causality are found in Brian Clegg, *Science for Life* (London: Icon Books, 2015), pp. 1–118.
257—Neil Turok's argument that physics is becoming too driven by mathematical theory is from M. Brooks, "Battle for the Universe," *New Scientist* (July, 2015), p. 6.
258—Isaac Asimov's comments on the unique nature of mathematics are from the foreword of Carl B. Boyer and Uta C. Merzbach, *A History of Mathematics* (New York: John Wiley & Sons, 1991), p. vii.
260—Richard Hamming's observation that we select the mathematics to fit the situation is from Richard Hamming, "The Unreasonable Effectiveness of Mathematics," *American Mathematical Monthly* 87, no. 2 (February 1980), pp. 81–90.

1. 本部分所标页码都为原版书页码。——编者注